The Role of Lower Airway Resonances in Defining Vowel Feature Contrasts

by

Steven Michael Lulich

Submitted to the Harvard-MIT Division of Health Sciences and Technology in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Speech and Hearing Bioscience and Technology

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2006

Author .....
Harvard-MIT Division of Health Sciences and Technology
September 1, 2006

Certified by .....
Kenneth N. Stevens
Clarence J. LeBel Professor of Electrical Engineering and Computer Science and Professor of Health Sciences and Technology
Thesis Supervisor

Accepted by .....
Martha L. Gray
Edward Hood Taplin Professor of Medical and Electrical Engineering
Co-Director, Harvard-MIT Division of Health Sciences and Technology

The Role of Lower Airway Resonances in Defining Vowel Feature Contrasts

by

Steven Michael Lulich

Submitted to the Harvard-MIT Division of Health Sciences and Technology
on September 1, 2006, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Speech and Hearing Bioscience and Technology

Abstract

Since the voicing source is located between the lower and upper airways and has a high impedance, the resonances of the lower airway appear as pole-zero pairs in vowel spectra. These pole-zero pairs interact non-linearly with the vocal tract formants, producing narrow frequency bands within which formant structure is unstable. The broader frequency bands between lower airway resonances are thus potentially optimal for the reliable production and accurate perception of specific formant patterns. This possibility is explored from three directions. First, models of the lower airway are built and analyzed, and their effects on vowel spectra are characterized. Second, evidence for the non-linear interactions between formants and lower airway resonances is presented from a speech production study, and the relations between these non-linearities and certain distinctive feature contrasts are explored. Third, a speech perception experiment is carried out in which the identification of a vowel (which could be either [+back] or [-back]) is dependent upon the interaction of the second lower airway resonance (represented as a zero without an accompanying nearby pole) with the second formant. The results of these studies indicate that lower airway resonances do play a role in speech production and perception, and that further study is warranted. In addition, some potential applications to respiratory and vocal medicine are suggested.

Thesis Supervisor: Kenneth N. Stevens
Title: Clarence J. LeBel Professor of Electrical Engineering and Computer Science and Professor of Health Sciences and Technology

Acknowledgments

As I sit down to write this acknowledgments section, I feel at a loss. Where to begin? I am afraid that it is impossible for me, in this tiny space, satisfactorily to thank and acknowledge all those who have made such positive impacts in my life and work. Nevertheless, the attempt must be made! My advisor, Ken Stevens, has been an incredible teacher and a joy to be around. I feel truly blessed for the opportunity to get to know him personally and to learn from him, and I know that I will be a much better teacher, researcher, and person because of him. I would also like to thank the other members of my dissertation committee: Adam Albright, David Gow, and Bob Hillman. I have benefitted tremendously from long dicussions with each of them, and they have put considerable time and energy into committee meetings and the reading of documents that have eventually wound up as part of this thesis. I am also thankful for the guidance I received from my PAC committee members and my oral exam committee members: Stefanie Shattuck-Hufnagel, Janet Slifka, Donca Steriade, David Gow, and Joe Perkell. I must pause to point out that Janet played a large role in helping me learn how to analyze electrical circuits, and was always patient with me and all my questions. I must also point out that Stefanie has been all-around wonderful - an ever ready source of encouragement, and frequently available for long evening discussions over pizza. Finally, I'd also like to thank Satra Ghosh for the time he has spent discussing lower airway resonances with me, and playing with real and synthetic speech to see what kinds of effects of the lower airway we could detect in analysis and perception.

I must acknowledge a few more faculty members whose influence will not be as readily apparent in these pages as I had hoped. Nevertheless, Sylvain Bromberger, Noam Chomsky, Morris Halle, and Nelson Kiang have played a particularly large role in helping me to view the bigger, broader picture.

I am also grateful for the wonderful camaraderie that I have enjoyed with the other students in the SHBT program, in the Speech Group, and in a few other places. First I must acknowledge Asaf Bachrach and Nick Malyska, with whom my adventures into the lower airway really began, and with whom some of those adventures continue. I must also point out Tamas Bohm, Xuemin Chi, Xiaomin Mou, Tony Okobi, and Morgan Sonderegger, who have in one way or another helped me to understand speech acoustics and language, and kept me sane in stressful times. To close the professional part of these acknowledgements, I am happy to thank my sister Rachel, who helped me in the last weeks to organize data as I was scrambling to get it ready for Chapter 4.

In less academic, but more personal ways, I am very thankful for my family - Mom, Dad, Ben, and Rachel - who have always been there for me and encouraged me to do my best. I am thankful for the friendship of Lindsay Whaley and Ioana Chitoran - my first teachers in linguistics. I am grateful to Lindsay and to Frank Li, who let me come with them to China for fieldwork that has unfortunately not been able to fit in this thesis - I hope the fieldwork will turn into some nice papers in the near to intermediate future. Finally, I would like to thank my church families for their support throughout the years, for the things I have learned from them, for the ways in which they have helped me become more and more the person that God wants me to be.

Soli Deo Gloria.

Contents

List of Figures

List of Tables

Preface

"The field is an attractive one; but those who would work in it need to be well equipped, both on the physical and on the phonetic side."
–J. W. S. Lord Rayleigh, *The Theory of Sound*.

In the past two centuries linguistics has undergone a number of major transformation. Earlier linguists (philologists) were concerned with the written word and the connections between and among ancient and modern written languages. More recently, linguists have come to focus ever more attention on problems of how humans communicate by means of language in real time. It is no longer enough to observe that /p/, /t/, and /k/ became /f/, /θ/, and /x/ in Germanic (Grimm's Law), for that observation says nothing about how people produce and perceive the sounds of language, neither today nor in the past. Distinctive feature theory developed at the cross-roads of two different kinds of linguistics. On the one hand, Roman Jakobson [29] and Nikolai Trubetzkoi [54] first introduced distinctive features within a structuralist framework, in order to systematically and concisely represent the collections of sounds (phonemes) that various languages use. On the other hand, Noam Chomsky and Morris Halle [9] discovered that distinctive feature theory had immediate application to problems in generative grammar. The shift from structuralism to generative grammar was only the latest step in the progression of linguistics from a branch of history to a branch of biology (as Chomsky puts it [8]). I do not claim

that linguistics has completed that progression.) Similarly, distinctive feature theory became grounded in the physics of the speech production and perception processes with Jakobson, Fant, and Halle's [30] explorations at about the same time.

Within the field of linguistics, distinctive feature theory is enormously important and successful. To a physiologist, this fact may carry little weight. The physiologist wants to know: Can distinctive features be measured? Do they have mass or energy? So far as I am aware, no linguist would answer these questions in the affirmative. Linguistics as such is a very indirect science. But I would maintain that the theory is important for physiologists. As we learn more about the physiology, physics, and acoustics of human speech communication, it is vital to keep in mind that we are studying a complicated system that has a definite structure. Distinctive feature theory tells us a lot about that structure, which we might otherwise loose sight of if we paid attention only to low-level measurable quantities with a definite mass or energy. The speech physiologist who ignores linguistics is in danger of missing the forest for the trees.

Nevertheless, the forest does contain trees, and the linguist must pay attention to the physiology and the acoustics of speech production and perception. In the past two centuries at least, linguistics has moved closer and closer to physiology. It is not yet a physiological science, but it is moving in that direction, and I believe it will continue in that direction. For the time being, I view distinctive feature theory as a powerful approximation to what is going on physiologically. It is only an approximation, and must be changed, refined, redefined, and made more concrete. The question remains how to get there. I think this goal is still quite a ways off on the horizon, but it should nevertheless be the goal.

Chapter 1

Introduction

1.1 Overview

The questions addressed in this thesis pertain to the role that the lower airway (below the glottis) plays in defining vowel contrasts in both speech production and speech perception. Because the impedance of the glottis, Z_g, is large, the influence of the lower airway on voiced vowel spectra is minimal. For this reason it has generally been assumed that the lower airway can be left out of models of vowel production and analyses of vowel acoustics. Indeed, both the modeling and analysis of vowels have generally been approached from the perspective that only the first few formants are important. Without doubt, the first few formants carry the greatest part of the burden in speech production and perception. However, the purpose of this thesis is to demonstrate that the lower airway also plays a significant role in both the production and perception of speech. The role of the lower airway is hypothesized to be of a different kind than the role of the vocal tract itself. While the vocal tract geometry is principally responsible for determining the frequencies of formants, the lower airway defines a set of frequency bands over which the formants range. It will be argued that

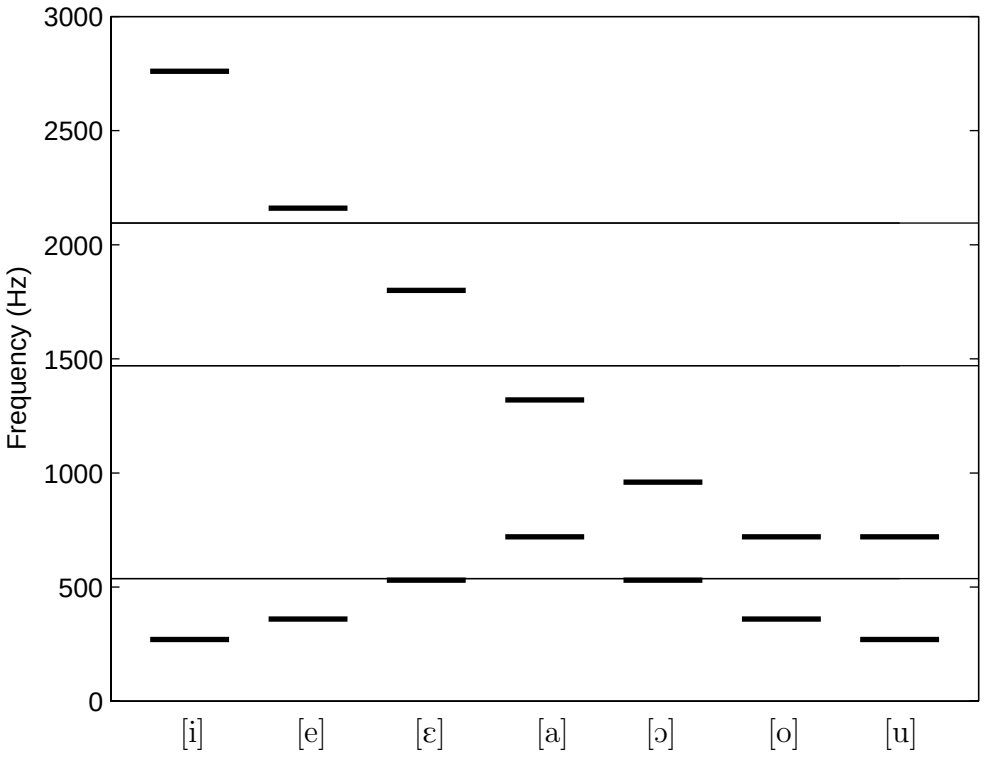


Figure 1-1: Frequency bands defined by the lower airway and vowel categories defined by these bands. The thin lines indicate the frequencies of the lower airway resonances; the thick lines indicate F1 and F2 frequencies.

these bands are important for defining vowel categories. For instance, a vowel of one category may have a formant in one band, while a minimally contrasting vowel may have the same formant in an adjacent band (cf. Figure 1-1).

The role of the lower airway in defining vowel contrasts is explored from three directions. First, models of the lower airway and the vocal tract are implemented in MATLAB. Details and results of the lower airway model are discussed in Chapter 2, while details and results of the combined vocal tract plus lower airway models are discussed in Chapter 3. The modeling work is intended to give a characterization of the acoustic effects of the lower airway on vowel spectra. Second, acoustic analyses

of naturally produced vowels are presented in Chapter 4. The purpose of this work is to show that the effects of the lower airway in speech production are frequently and clearly observable. Third, data from speech perception experiments are discussed in Chapter 5. The perceptual work is intended to demonstrate that the effects of the lower airway in speech acoustics are perceptually salient. The broader import of Chapters 2 through 5 is discussed in Chapter 6, and Chapter 7 summarizes the thesis work and concludes.

1.2 Background: The theoretical acoustic framework

A linear source-filter model of speech production will be assumed throughout this thesis. Because the research focuses on vowels, in which pressure in the pharyngeal and oral cavities is close to atmospheric, non-acoustic aerodynamic processes (such as aspiration and frication noise generation) and source-filter interactions will generally be assumed to be negligible.

The result of coupling the lower airway to the vocal tract in vowel production is, among other things, the introduction of a set of poles and zeros. For this reason, the standard all-pole model of vowel production does not hold. In an all-pole model, formants are the only characteristics of the vowel spectrum that need be specified [13]. In the model that will be developed here, a more precise terminology is required.

Poles and zeros, strictly defined, are frequency-bandwidth ordered pairs in the complex s-plane. The poles and zeros that will be discussed in this thesis all lie in the left half-plane. Poles and zeros lying off the real axis are always paired with their complex conjugates. The transfer function for a system with a complex conjugate pair of poles (or zeros) is determined using the method described in Fant [13] (cf.

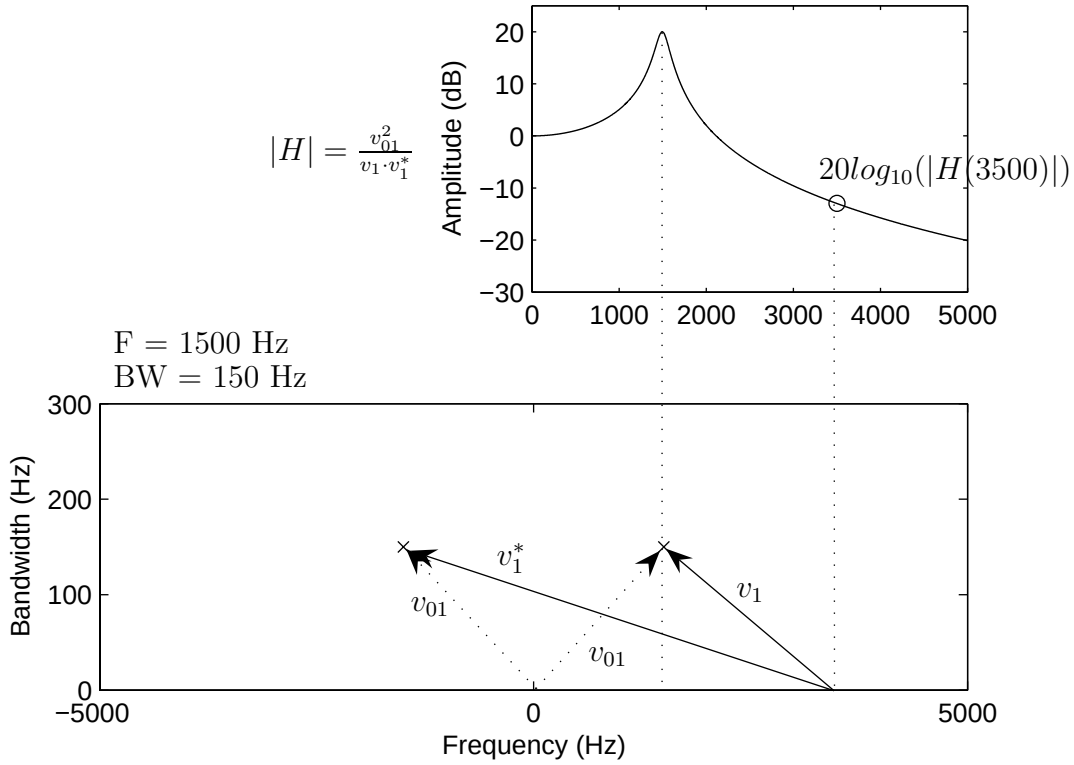


Figure 1-2: A complex conjugate pair of poles and its spectrum.

Figure 1-2). At the risk of some ambiguity, each such spectrum will itself be referred to as a pole (or zero) in this thesis. In an all-pole model, poles are the formants. In the model developed here, the formants will be defined more restrictively (see below).

In an acoustic system, the poles are defined by the entire system. Certain parts of the system, however, if they are weakly coupled to the rest of the system, may contribute a greater or lesser degree of influence on particular poles. For example, if a part of a larger system consists of a single uniform tube weakly coupled to the rest of the system, a certain set of poles will be defined primarily by the simple tube (at frequencies close to $\frac{nc}{2l}$ or $\frac{(2n-1)c}{4l}$, where n is an integer, l is the length of the tube, and c is the speed of sound), while other poles will be defined primarily by the other

parts of the system. Since in the speech production system the glottal impedance is large, the vocal tract and the lower airway are weakly coupled, so that some of the whole system's poles are defined primarily by the vocal tract while others are defined primarily by the lower airway. The poles defined primarily by the vocal tract are the poles which we will refer to as formants.

Because the lower airway is below the location of the glottal phonation source, it contributes both poles and zeros to the vowel spectrum, whereas the vocal tract itself contributes only poles (assuming the absence of side branches such as the nasal and sublingual cavities). The poles and zeros contributed by the lower airway are paired together, and will be referred to as pole-zero pairs. In the extreme case that the glottal impedance is infinite (hence vocal tract and lower airway coupling is infinitessimal), the lower airway pole-zero pairs occupy the same locations in the s-plane and cancel each other out. This is mathematically equivalent to a model of the vocal tract which does not include the lower airway. As the glottal impedance is decreased (made finite), the lower airway poles and zeros separate; the poles, influenced by the geometry of the vocal tract above the glottis, shift in frequency and bandwidth while the zeros, wholly defined by the lower airway, remain in place.

1.3 Background: The theoretical linguistic framework

Generative linguistics and specifically generative phonology will be assumed throughout the thesis. The phonological framework employed will make extensive use of distinctive features, along the lines indicated in Bromberger and Halle [2]. The phonetic framework employed is based on the quantal theory [46, 47, 48]. The goal of this thesis is to test to what extent lower airway resonances define quantal boundaries

between vowels differing by a single distinctive feature.

1.4 Background for chapter 2: Models of the lower airway

Studies of the lower airway acoustic system have come largely from two disciplines: speech science and respiratory science. In the speech science literature there are four noteworthy papers: by van den Berg [55], Fant et al. [14], Ishizaka et al. [26], and Cranen and Boves [10]. Van den Berg [55] constructed an electrical analog model of the subglottal system and compared its behavior to impedance measurements made with human and canine cadavers. Van den Berg is well known for championing the myoelastic-aerodynamic theory of voice production, and it was largely for the purpose of developing this theory that the subglottal system attracted his attention, for the presence of standing waves below the glottis will affect the vibration of the glottis during phonation (cf. later work by Flanagan [16], and Fant [1]). The same issue motivated the studies by Ishizaka et al. [26] and Cranen and Boves [10]. Ishizaka et al. [26] developed their own electrical analog model and made impedance measurements at the tracheostoma of laryngectomized Japanese patients. Cranen and Boves [10] constructed a very simple model based on the results of Ishizaka et al. [26] and analyzed pressure signals measured just below the glottis. Unlike van den Berg [55] or Ishizaka et al. [26], Cranen and Boves [10] linked their findings directly to vocal fold vibration, and showed that certain assumptions about the closed phase of glottal vibration were unfounded. Fant et al. [14] were interested more in the spectral components of speech signals filtered by the vocal tract, rather than in the nature of the source spectrum itself. However, their study focused on the contribution of subglottal resonances to the speech signal in aspirated sounds. During aspiration, the glottal

area is larger than in voiced speech, and therefore the glottal impedance is reduced, allowing for greater effects of the subglottal system on the speech signal. Although they made reference to a possible effect of the subglottal system in voiced speech, they did not publish any further research in this direction. The most recent studies of the subglottal system by speech scientists [3, 5] analyzed data from accelerometers placed on the skin of the neck just below the level of the glottis. Cheyne's [3] study was motivated by the possibility of detecting voice pathologies using subglottal acoustic signals. Chi and Sonderegger [5], following Stevens' [48, pp. 299-303] suggestion, performed the first study in which the effect of subglottal resonances on the vocal tract transfer function in voiced speech was explored.

In the respiratory science literature there are more studies of the subglottal acoustic system. Several of these studies are concerned with the input impedance of the lower airway up to only several tens of cycles per second. Four papers of note concerned with input impedances up to frequencies relevant for speech were published by Fredberg and Hoenig [17], Hudde and Slatky [25], Habib et al. [19], and Harper et al. [22]. These studies have invariably focused on the use of acoustic signatures of the lower airway to aid in the detection and diagnosis of lung pathologies, although no clinical applications have yet resulted from this work.

Most of the speech and respiratory studies presented above made use of similar simplifying assumptions. Notably, most of them assumed that the bronchial tree branches symmetrically at all levels, and most of them relied on statistical averages of the dimensions of the trachea, bronchi, and lobuli as published in the respiratory literature (chiefly from Rohrer [44], and Weibel [57]). There are some exceptions. Fredberg and Hoenig [17] constructed a model with asymmetrical branching, but constrained so that every bifurcation at a given level gave rise to the same pair of (asymmetric) daughter branches, so that the overall model was symmetric about the main tracheal axis. Hudde and Slatky [25] constructed a model of one half of the

lung with asymmetrical branching, in which the dimensions of the daughter branches were largely statistically determined, within reasonable bounds. Ishizaka et al. [26] uniformly scaled down the Western length dimensions given in Weibel [57] in order to accommodate the relatively smaller size of Japanese speakers. Habib et al. [19] uniformly scaled the length and diameter dimensions given by Horsfield et al. [24] to fit their model to seven individuals whose subglottal impedance characteristics were measured by means of an endotracheal tube.

A detailed model of the lower airway has never been constructed for use in determining the effects of the lower airway resonances in vowel production. The construction and study of such a model can lead to new ways of analyzing and interpreting vowel spectra and spectrograms, voice quality, and lung health.

1.5 Background for chapter 3: Models of the vocal tract

Although many researchers have constructed and studied models of he vocal tract, few have studied vocal tract models coupled with a lower airway. When such studies have been carried out, it has always been with a maximally simplified model of the lower airway (usually a single tube ending in a large compliance) [45, 10, **?**]. In this thesis, the model of the lower airway described in Chapter 2 will be coupled with a model of the vocal tract in order to study the acoustic effects of this coupling.

1.6 Background for chapter 4: Vowel acoustics

The phonetic study of vowels has undergone a steady progression and sophistication since Willis' [58] seminal study. One way to characterize the progression is as follows:

First, vowels were characterized as a single 'cavity tone' excited by a 'larynx tone'. Later, it was found that two cavity tones were characteristic of vowels [6]. Finally, with the development of more precise measurement and analysis methods, the acoustic theory of speech production [13] was developed, in which a large number of formants are recognized in each vowel, only the lowest two or three of which are primarily responsible for defining the vowels. In this thesis, the need to consider a still more detailed perspective on vowel acoustics will be demonstrated. Specifically, it will be shown that in addition to formants there are pole-zero pairs arising from vocal tract-lower airway coupling which play a significant role in defining vowels. The effects of these pole-zero pairs (e.g. additional poles in vowel spectra, discontinuities in formant tracks) are frequently and clearly observable, if one looks for them.

Chapter 2

Modeling the Lower Airway

2.1 Implementation of the model

A model of the lower airway was implemented in MATLAB. Data for the dimensions of the first 13 generations of the bronchial tree were taken from Horsfield et al. [24] (as reported in Harper et al. [23]; cf. Table 1), where the trachea is 0th generation, the two major bronchii are 1st generation, and so on. The length data in Figure 2-1 were scaled up by 10% in the model calculations in order to create a better match between the calculated impedance structure and the resonances measured by other investigators. This led to a modeled trachea 11 cm long, rather than 10 cm. This was justified, however, given that both van den Berg [55] and Weibel [57] reported tracheas with lengths greater than 12 cm.

The model is asymmetric, following human morphometric data, and it is assumed that every branching is binary. Because of the asymmetry, calculation of the total input impedance from the top of the trachea is complicated and requires a large amount of computer memory. For this reason, only the first 13 generations were included in the calculation. This is justified, however, because in the range 0 - 3000

Hz the input impedance of the system converges quickly as the number of generations is increased to 13 (see below).

It is assumed that every airway branches into two smaller airways: there is 1 member of the 0th generation of airways - the trachea; there are 2 members of the 1st generation; there are 4 members of the 2nd generation; there are 8 members of the 3rd generation; and so on. The number of airways in the n^{th} generation is therefore 2^n.

In the model, each airway is considered to be a soft-walled cylindrical tube with constant cross-sectional area. Ignoring the effect of the walls for the moment, the input impedance at the top of an airway that branches into two airways is given by Equation 2.1, which is derived from Morse's [43] equation 24.17.

$$Z = \frac{\rho c}{A}\left[\frac{Z_L + j\frac{\rho c}{A}\tan(kl_1)}{\frac{\rho c}{A} + jZ_L\tan(kl_1)}\right] \tag{2.1}$$

Z_L is the parallel impedance of the two daughter airways, ρ is the density of air, c is the speed of sound, A is the cross-sectional area of the parent airway, l_1 is the length of the parent airway, and $k = \omega/c$ is the wave number.

The input impedance at the top of an airway therefore depends upon the input impedances of its two daughter airways. In order to determine the input impedance at the top of the trachea, then, the number of generations to include in the calculation must first be determined, as well as the load impedance that will be assumed for the most peripheral branches. However, the peripheral load impedance can be neglected when the number of generations is large (as illustrated in Figure 2-2; cf. [28]).

Because the number of airways in the n^{th} generation is 2^n, the total number of airway impedances to be calculated is $2^{n+1} - 1$. If n = 13, the total number of airways is $2^{13+1} - 1 = 16383$. The impedance of each of the 8192 airways of the 13th generation must be calculated, and then pairwise added in parallel. The resulting

Table 2.1: Morphometric data used in the model, taken from [24].

Depth, n	Tube length, l [cm]	Tube radius, r [cm]	Wall thickness, h	Fraction of cartilage, c_{frac}	Recursion Index, $\Delta(n)$
0	10.0	0.80	0.3724	0.76	1
1	5.0	0.6	0.1735	0.5000	2
2	2.2	0.55	0.1348	0.5000	3
3	1.1	0.40	0.0528	0.3300	3
4	1.05	0.365	0.0409	0.2500	3
5	1.13	0.295	0.0182	0.2000	3
6	1.13	0.295	0.0182	0.0922	3
7	0.97	0.270	0.0168	0.0848	3
8	1.08	0.215	0.0137	0.0669	3
9	0.950	0.175	0.0114	0.0525	3
10	0.860	0.175	0.0114	0.0525	3
11	0.990	0.155	0.0103	0.0449	3
12	0.800	0.145	0.0097	0.0409	3
13	0.920	0.140	0.0094	0.0389	3
14	0.820	0.135	0.0091	0.0369	3
15	0.810	0.125	0.0086	0.0329	3
16	0.770	0.120	0.0083	0.0308	3
17	0.640	0.109	0.0077	0.0262	3
18	0.630	0.100	0.0072	0.0224	3
19	0.517	0.090	0.0066	0.0000	3
20	0.480	0.080	0.0060	0.0000	3
21	0.420	0.070	0.0055	0.0000	3
22	0.360	0.055	0.0047	0.0000	2
23	0.310	0.048	0.0043	0.0000	2
24	0.250	0.038	0.0038	0.0000	1
25	0.110	0.032	0.0034	0.0000	0
26	0.131	0.027	0.0032	0.0000	0
27	0.105	0.024	0.0031	0.0000	0
28	0.075	0.022	0.0030	0.0000	0
29	0.059	0.040	0.0039	0.0000	0
30	0.048	0.040	0.0039	0.0000	0
31	0.048	0.040	0.0039	0.0000	0
32	0.048	0.040	0.0039	0.0000	0
33	0.048	0.040	0.0039	0.0000	0
34	0.048	0.040	0.0039	0.0000	0
35	0.048	0.040	0.0039	0.0000	0

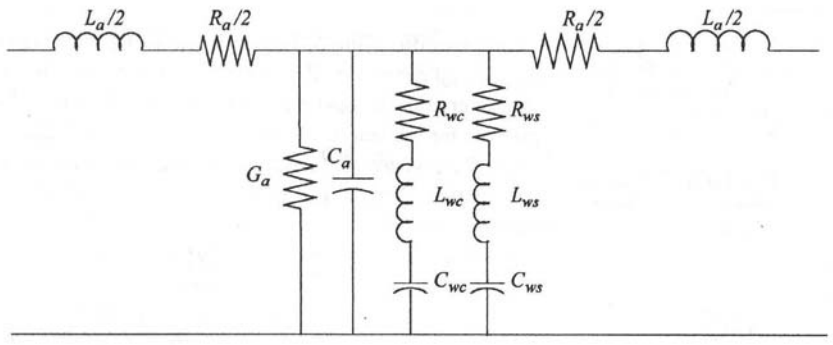


Figure 2-1: T-network (adapted from Harper et al. [23]). Lumped element values are given in Tables 2.2 2.5.

set of 4096 impedances form the load impedance for each of the 4096 airways of the 12th generation. The cycle is repeated until the input impedance from the top of the trachea is known.

Because the walls in the model are not rigid, the implementation of the model is slightly more complicated and Equation 2.1 cannot be used. Instead, each airway was modeled as a transmission line composed of a series of T-networks, as in Figure 2-1.

Each T-network represents 1.5 cm of length or less. Each airway is modeled, therefore, as a series of 1.5 cm T-networks plus one or zero residual T-networks of length smaller than 1.5 cm. A basic length of 1.5 cm ensures reasonable accuracy up to greater than 3000 Hz.

Data for the wall properties were taken from Harper et al. [23] (cf. Table 2.1). The model was built with an option to use rigid walls in order to compare the lower airway input impedances with and without shunting elements due to wall properties. Equations used to compute the values of the lumped elements of the transmission line were also taken from Harper et al. [23] and are given in Tables 2.2 through 2.5. (See Appendix A for the MATLAB implementation of the lower airway model.)

Table 2.2: Acoustic lumped elements.

Parameter	Value	Units
Resistance	$R_a = \frac{2l}{\pi r^3}\sqrt{\frac{\omega\rho_0\eta}{2}}$	$\frac{dyne\cdot s}{cm^5}$
Inertance	$L_a = \frac{\rho_0 l}{A}$	$\frac{dyne\cdot s^2}{cm^5}$
Compliance	$C_a = \frac{Al}{\rho_0 c^2}$	$\frac{cm^5}{dyne}$
Conductance	$G_a = 2\pi r l \frac{\nu-1}{\rho_0 c^2}\sqrt{\frac{\kappa\omega}{2c_p\rho_0}}$	$\frac{cm^5}{dyne\cdot s}$

Table 2.3: Properties of air.

Property	Air
$\rho_0(g/cm^3)$	$1.14 \cdot 10^{-3}$ (moist air, 37^oC)
$\eta(dyne \cdot s/cm^2)$	$1.86 \cdot 10^{-4}$ (20^0C, 1 atm)
$\nu = c_p/c_\nu$	1.4
$\kappa(cal/cm - s - {}^0C)$	$0.064 \cdot 10^{-3}$ (37^0C)
$c_p(cal/g - {}^0C)$	0.24 (0^0C, 1 atm)
$c(cm/s)$	$3.54 \cdot 10^4$ (moist air, 37^0C)

Table 2.4: Non-rigid wall lumped elements.

Parameter	Value	Unites
Resistance	$R_{wxt}(\omega) = \frac{\eta_{wx}(\omega)h}{2\pi r^3 l}$	$\frac{dyne\cdot s}{cm^5}$
Inertance	$L_{wxt}(\omega) = \frac{\rho_{wx}h}{2\pi r l}$	$\frac{dyne\cdot s^2}{cm^5}$
Compliance	$C_{wx}(\omega) = \frac{2\pi r^3 l}{E_{wx}(\omega)h}$	$\frac{cm^5}{dyne}$

Table 2.5: Properties of non-rigid walls.

Parameter	Default Value
Soft Tissue Density (ρ_{ws})	$1.06g/cm^3$
Soft Tissue Viscosity (η_{ws})	$1.6 \cdot 10^3 dyne \cdot s/cm^2$
Soft Tissue Elasticity (E_{ws})	$0.392 \cdot 10^6 dyne/cm^2$
Cartilage Density (ρ_{wc})	$1.14g/cm^3$
Cartilage Viscosity (η_{wc})	$180.0 \cdot 10^3 dyne \cdot s/cm^2$
Cartilage Elasticity (E_{wc})	$44.0 \cdot 10^6 dyne/cm^2$

Once the input impedance of the lower airway has been calculated, it can be incorporated into a model of the vocal tract, and its influence on vowel spectra (transfer functions from the glottis to the lips) can be studied. This will be discussed in the next chapter.

2.2 Results

2.2.1 *Comparison of asymmetric and symmetric models*

Figure 2-2 shows the frequencies and amplitudes (in dB) of the poles and zeros of the lower airway input impedances for the asymmetric model and a symmetric model as a function of the number of branching generations included in the calculation. For each case the load impedance at the periphery is $Z_L = 0$. Thus, for the zero generation only the trachea is modeled, with an open end at the periphery, yielding a lowest pole with quarter-wavelength frequency $P1 = \frac{c}{4l} = \frac{35400cm/s}{4*11cm} \approx 805Hz$. As the number of generations increases, the poles and zeros decrease in frequency. The amplitudes of the poles decrease while the amplitudes of the zeros increase. Figure 2-3 illustrates this for the asymmetric model with 2, 4, and 8 generations.

The fact that the frequencies and amplitudes of the input impedance poles and zeros converge with increasing number of generations is an indication that the peripheral load impedance can take on any value without significantly altering the input impedance. This was noted before by Jackson et al. [28] and Hudde and Slatky [25], and results from the rapidly increasing total cross-sectional area as each new generation increases the number of peripheral airways by a factor of 2. Figure 2-4 gives a rough estimate of this increase, as follows.

For each airway in the asymmetric model, one daughter branch is shorter than the other. Calculating the total cross-sectional area for the asymmetric model is

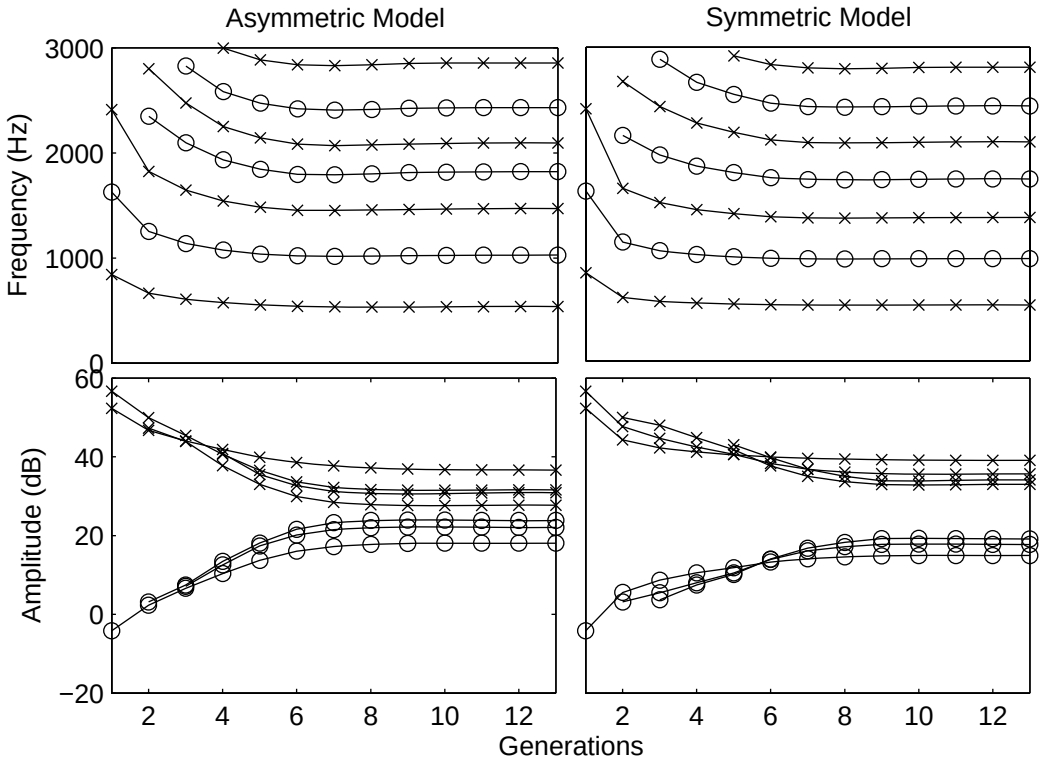


Figure 2-2: Frequencies and amplitudes of the lower airway input impedance poles and zeros as the number of branching generations included in the model is increased.

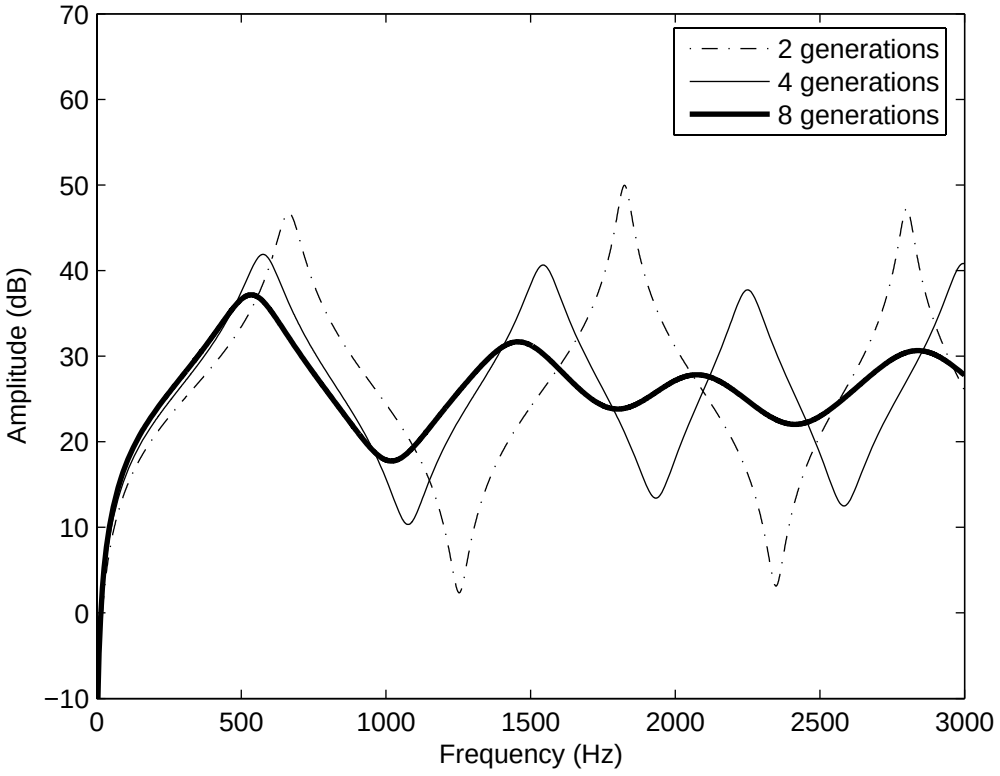


Figure 2-3: Lower airway input impedances for 2, 4, and 8 generations.

therefore difficult. We may, however, define its upper and lower limits by calculating the total cross-sectional areas of two symmetric models. In one symmetric model, the daughter branches are identical to the shorter branch of the asymmetric model; in the other symmetric model, the daughter branches are identical to the longer branch of the asymmetric model. Since the length and the radius of airways correlate with each other, the total cross-sectional area of the longer symmetric model will become large more quickly than the total cross-sectional area of the shorter symmetric model. The total cross-sectional area of the asymmetric model lies somewhere in between. In Figure 2-4, the solid lines show the total cross-sectional areas for the symmetrical models, and the dashed line shows the data from Weibel [57, p. 139]. Weibel's [57] data lie close to the line marking the lower limit for the model calculations. It it therefore likely that the total cross-sectional area of the asymmetric model will grow more quickly than Weibel's [57] data.

The upper limit on the total cross-sectional area corresponds to the symmetric model whose input impedance poles and zeroes are plotted in Figure 2-2. Both the symmetric and asymmetric models converge after 6 generations, and the frequencies of their poles and zeros are similar. However, the poles and zeros are slightly more evenly spread in the symmetric model than in the asymmetric model, and they are more prominent in the symmetric model.

The input impedance of the asymmetric model calculated with 13 generations is given in Figure 2-5. The solid line is the input impedance with peripheral load impedance $Z_L = 0$, and the dashed line is the input impedance with peripheral load impedance $Z_L \to \infty$. The inset shows the difference between the zero and infinite load impedance conditions, which is smaller than 1 dB for all frequencies.

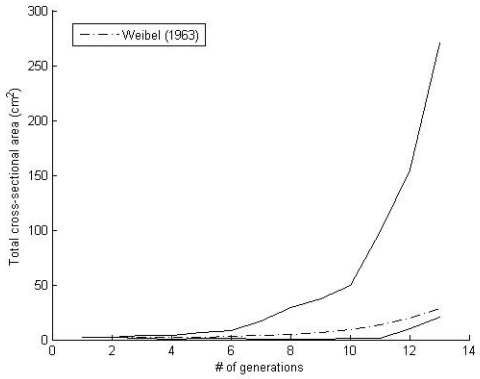


Figure 2-4: Total cross-sectional area of the lower airway as more generations are added.

2.2.2 *Evaluation of the asymmetric model*

In Figure 2-6, previously reported lower airway impedance data are plotted versus the predictions of the asymmetric model (see also Tables 2.7, 2.8, and 2.9). In the left-hand column of the Figure (a, c, e, g), data were drawn from papers in which the measurements of the lower airway input impedance poles and zeros were made indirectly by analyzing the speech signal. In the right-hand column, data were drawn from papers in which the measurements were made directly by various means (e.g. intubation or via tracheostoma) (b, d, f), or in which symmetric models of the lower airway (except van den Berg [55], who reported frequencies of the lower airway poles and zeros significantly lower in frequency than others [26]; see Tables 2.10 and 2.11) were used to calculate the poles and zeros directly (h). Data in (a), (b), and (e) were from male subjects, and data in (c), (d), and (g) were from female subjects. Data in (f) were group averages, mostly of male subjects. Poles are indicated by 'x' and zeros by 'o'. The x-coordinate of each point is the frequency of that pole or zero as calculated by the asymmetric model described above. The y-coordinate of each point is the frequency of that pole or zero as reported in the literature. The diagonal solid

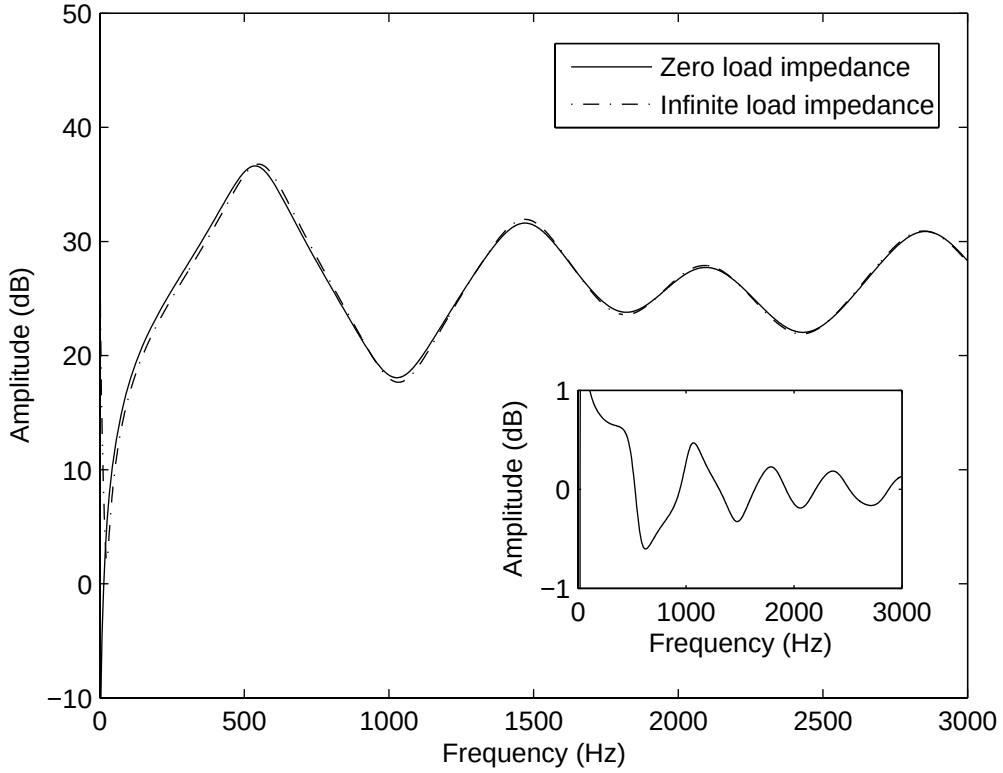


Figure 2-5: Lower airway input impedance when 13 generations are modeled. The solid line indicates a peripheral load impedance $Z_L = 0$; the dashed line indicates a peripheral load impedance $Z_L \to \infty$. The inset indicates the difference between the two load impedance conditions.

lines have slope $m = 1$, so that the more tightly the data cluster around this diagonal, the better they are predicted by the model. The diagonal dashed lines are the linear regression lines which best fit the data (in terms of least squared errors).

Since the model was based on male anatomical data, it is more likely to fit the data measured directly from male subjects (b), or the group averages of the same (f). That this is indeed the case is readily seen in the plots themselves (see Table 2.6 for some statistical measures of the regression fit for each of the plots). For the data measured directly from female subjects (d), the values predicted by the model are slightly too low, consistent with the use of larger airway dimensions. The data measured indirectly for male and female subjects (a, c) are more difficult to interpret in terms of the model predictions. This is due less to the larger spread of the data than to the fact that poles of the input impedance appear as zeros in the speech signal, and poles in the speech signal appear between adjacent poles and zeros in the input impedance (the pole in the speech signal is always at a higher frequency than the zero in the speech signal). Since the poles of the input impedance correspond more or less exactly to the zeros in the speech signal, they are replotted as such in (e) and (g). With the exception of a raised frequency of the first pole, these graphs are similar to those in (b) and (d), as expected. Similar to what was found above, frequency data from symmetric models (h) are similar to the predictions of the asymmetric model.

The Q values of the poles and zeros of the lower airway input impedance modeled using a symmetric and an asymmetric lower airway are give in Table 2.12, along with Q values reported in the literature (except for van den Berg [55], for which see Tables 2.10 and 2.11).

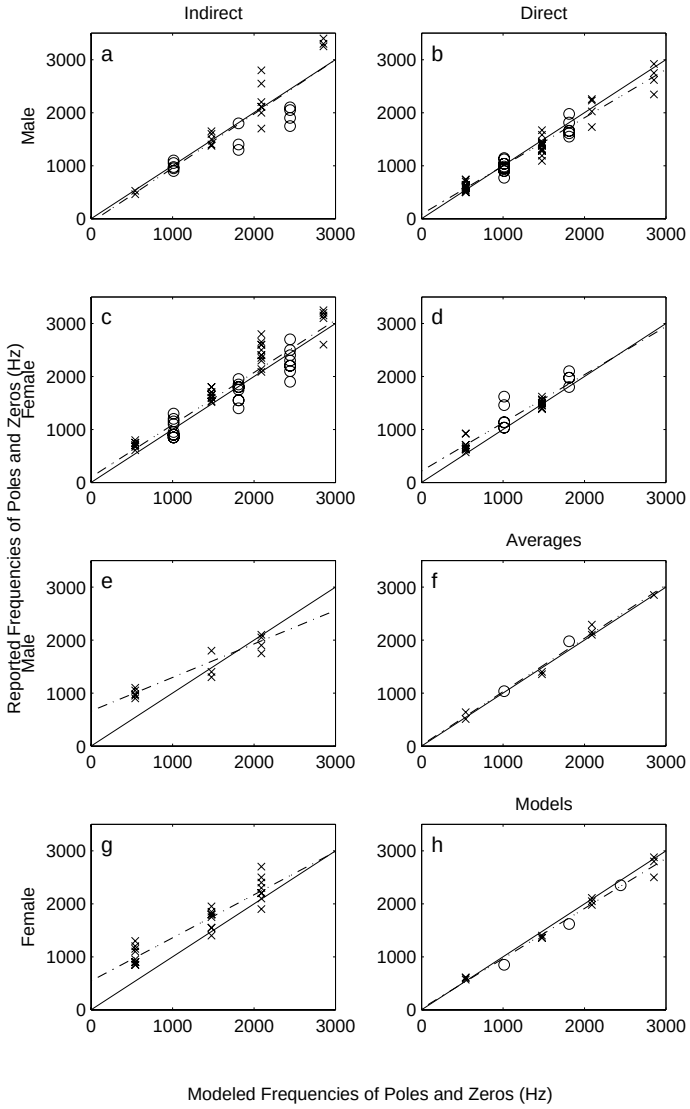


Figure 2-6: Comparison of the modeled lower airway input impedance pole 'x' and zero 'o' frequencies with those reported in the literature. In the first and third rows, data are reported from male subjects; in the second row and in (g), data are reported from female subjects; in (h), data are reported from models. In the right column, data are reported in which direct (e.g. input impedance) measurements were made (or modeled); in the left column, data are reported in which indirect speech spectral measurements were made. The zeros in (a) and (c) are replotted as poles in (e) and (g) because the poles of the input impedance are zeros in the speech signal (see Chapter 3).

Table 2.6: Statistics of the fit between the model and the reported input impedance pole and zero frequencies. The root-mean-squared error and r^2 values refer to the regression line as the best fit for the data, for which the slope and y-intercept are given.

Type	y-intercept	slope	rms error (Hz)	r^2
male direct (b)	84.1	0.91	138	0.94
female direct (d)	216.7	0.91	153	0.89
averages (f)	17.2	1.01	104	0.98
models (h)	21.6	0.94	106	0.98
male indirect (a)	-66.1	1.02	319	0.81
female indirect (c)	100.3	0.99	245	0.88
male indirect replotted (e)	661.4	0.63	167	0.87
female indirect replotted (g)	545.0	0.81	182	0.90

2.2.3 *Applications to medicine*

The possibility exists that a model such as the one described in this chapter could be used to study the cavity affiliations of the input impedance poles and zeros. For instance, the second lower airway pole is largely affiliated with the left bronchial tree, as illustrated in Figure 2-7. If the right bronchial tree is removed and replaced with an infinite impedance, the frequency and amplitude of the second pole of the modified input impedance is similar to the frequency and amplitude of the second pole of the whole input impedance. On the other hand, if the left bronchial tree is replaced with an infinite impedance, the frequency and amplitude of the second pole of the modified input impedance is quite different from the frequency and amplitude of the second pole of the whole input impedance. Thus, if a patient presents with an abnormally low or high second lower airway pole, a physician might conclude that this patient might have lung disease localized in the left bronchial tree.

Of course, this example is rather crude, and a patient with lung disease advanced enough to diagnose by means of such an acoustic test might have other more obvious indications already. It is also possible that the complex geometry of the lower airway

Table 2.7: Data reported in the literature in which direct (e.g. input impedance) measurements were made.

Gender	Source	P1	P2	P3	P4	P5	Z1	Z2	Z3	Z4	Z5
male	CH[3]	640	1400				1035	1975			
		640	1400				900	1975			
		500	1400				1035	1975			
		550	1400				900	1975			
		500	1400				1035	1975			
	HA[22]	606	1279	2256	2919						
		529	1092	1729	2343						
		743	1191	2024	2617						
		699	1318	2228	2748						
	HR[19]	590	1392				984	1662			
		508	1284				774	1550			
		602	1428				1120	1654			
		602	1430				956	1610			
		636	1584				935	1814			
		575	1440				1146				
		732	1670				955				
	MS[45]		1405								
			1384								
			1303								
			1454								
	XC[4]	577	1374								
		625	1280								
		615	1310								
female	CH[3]	670	1550				1130	2100			
		640	1400				1035	1975			
		620	1500				1140	1800			
		570	1400				1035	1975			
		640	1400				1035	1975			
	HR[19]	920					1458				
		926					1618				
	MS[45]		1568								
			1496								
			1383								
			1508								
	XC[4]	708	1620								
		625	1469								
		708	1447								

Table 2.8: Data reported in the literature in which indirect speech spectral measurements were made.

Gender	Source	P1	P2	P3	P4	P5	Z1	Z2	Z3	Z4	Z5
male	CH[3]						1100				
							950				
							900				
							975				
							1050				
	HH[21]	528		2115							
		461	1376	2112							
	KK[37]		1500	2000				1300	1750	2400	
			1500		3400			1800		3200	
			1400	1700	3250				2100	3050	
			1650	2550	3300			1800	2050	3000	
			1600	2200				1800	2050		
			1600	2800				1400	1900	2400	
female	CH[3]						1100				
							1150				
							1150				
							1200				
							1300				
	HH[21]	607	1534	2139							
		671	1515	2086							
	KK[37]	750	1800	2650	3150		900	1550	2100	2900	
		700	1650	2500			850	1800	2100	3200	
			1700	2600			900	1850	2200		
		750	1600		3100		850	1800	2300		
			1650	2350			900	1950	2400		
		700	1650	2400	3250		850	1800	2200	3050	
		750					900				
			1800	2600				1550	2500	3200	
		800	1650	2300	2600		950	1750	2700		
			1700	2400			900		2200	3100	
			1600	2800	3200			1400	1900	2400	

Table 2.9: Data reported in the literature in which direct measurements (group averaged from male subjects) or models were made.

Type	Source	P1	P2	P3	P4	P5	Z1	Z2	Z3	Z4	Z5
averages	CB[10]	510	1355	2290							
	FF[14]	640	1400	2150	2850						
	IMK[26]	640	1400	2100			1035	1975			
models	FF[14]	600	1400	2050	2800	3450	850	1620	2350	3050	3700
	HA[22]	569	1360	1980	2500						
	IMK[26]	615	1355	2110	2879						

Table 2.10: Frequencies and Q values for the modeled lower airway input impedance poles and zeros reported by van den Berg [55].

	P1	P2	P3	P4	P5	P6
f	314	890	1390	1860	2415	2950
Q	7.1	15.9	12.6	16.9	25.4	28.1
	Z1	Z2	Z3	Z4	Z5	Z6
f	33.3	615	1175	1630	2140	2715
Q	2.4	11.6	13.1	14.8	26.8	33.9

Table 2.11: Frequencies and Q values for the lower airway input impedance poles and zeros (of a large dog) reported by van den Berg [55].

	P1	P2	P3	P4	P5	P6
f	302	870	1425	1700	2500	3050
Q	2.5	5.8		3.8	12.5	12.2
	Z1	Z2	Z3	Z4	Z5	Z6
f	40	740	1100	1500	2200	2750
Q	1.1	4.2	3.7		7.3	13.8

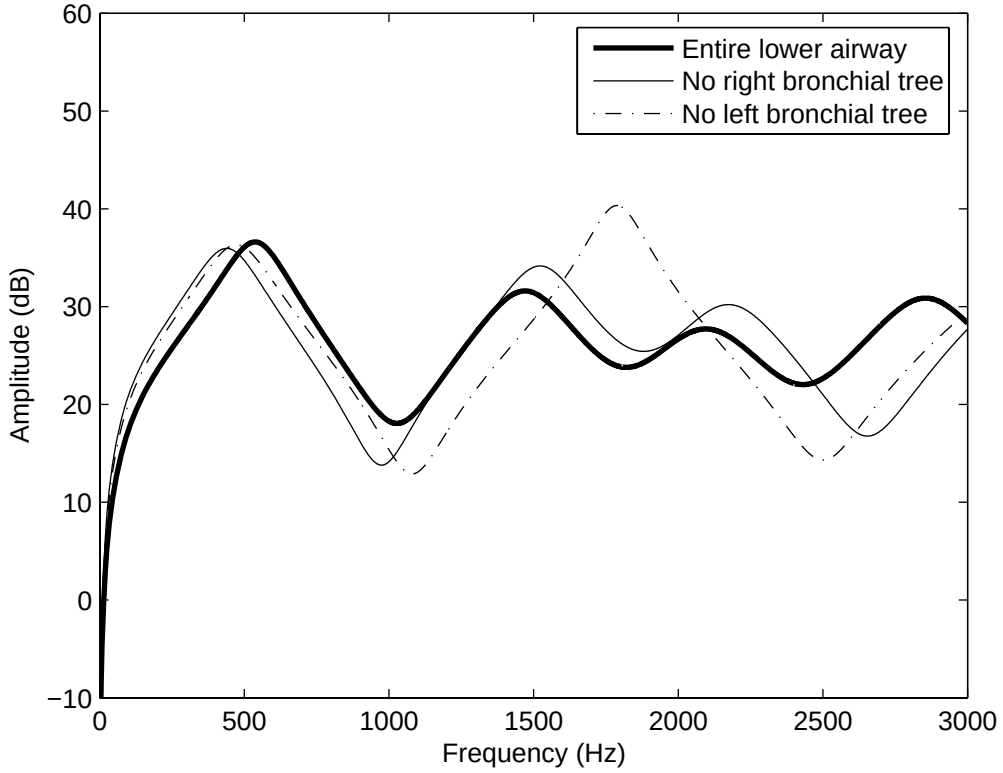


Figure 2-7: Lower airway input impedance dependence on gross lung structure. If the right bronchial tree is removed, the effect on the input impedance below 2500 Hz is minimal; if the left bronchial tree is removed, the effect on the input impedance above 1500 Hz is significant.

Table 2.12: Frequencies and Q values for the modeled lower airway input impedance poles and zeros, and Q values reported in the literature (except [55], for which see Tables 2.10 and 2.11). Cheyne [3] reported bandwidths for his 10 subjects individually, though they have been averaged here.

	P1	P2	P3	P4	Z1	Z2	Z3
asymmetric model	537	1470	2095	2855	1028	1821	2431
symmetric model	536	1372	2094	2807	979	1740	2438
	Q_{P1}	Q_{P2}	Q_{P3}	Q_{P4}	Q_{Z1}	Q_{Z2}	Q_{Z3}
asymmetric model	2.7	5.4	5.8	9.3	4.5	5.4	7.2
symmetric model	3.8	8.9	11.1	9.8	7.0	10.2	12.5
IMK[26] (modeled)	2.5	8.7	15.0	16.9			
CB[10] (measured)	4.9	8.8	6.4				
CH[3] (measured)	2.6	3.9			5.7	7.5	

will render more sensitive acoustic tests uninterpretable, for instance if the second pole of the input impedance is not strongly affiliated with any one region of the left bronchial tree. Nonetheless, the possibility that acoustic tests of lung health could be developed represents a strong warrant for continued research in this direction. If the acoustic input impedance structure can be correlated with lung structure in a rather straight-forward way, and if the relation between this input impedance structure and the pole-zero structure of speech signals is better understood, it may be possible to develop non-invasive, inexpensive, and quick and easy methods for detecting and diagnosing lung diseases. Further applications to medicine might be made from more detailed study of the effects of lower airway wall properties, which can have large effects on output speech spectra.

The relation between the lower airway input impedance structure and the pole-zero structure of speech signals will be investigated in the next chapter.

2.2.4 *Major conclusions of Chapter 2*

An asymmetric model of the lower airway was constructed based on human lung morphometric data reported by Horsfield et al. [24]. A symmetric model was also constructed and compared with the asymmetric model. For the frequency range 0 - 3000 Hz, it is sufficient to model a small number of generations of the branching bronchial tree (less than 8). The pole and zero frequencies of the lower airway input impedance obtained from the model fit well with measured data reported in the literature. On account of this fit, and because lung volume changes over the course of an utterance do not appear to affect the first few lower airway resonances [3], it appears justified to use this model in Chapters 3 and 4.

Chapter 3

Modeling the Effects of the Lower Airway on Vowel Spectra

3.1 Implementation of the model

In order to calculate the spectra of vowels with varying degrees of glottal coupling to the lower airway, the basic model of the vocal tract as shown in Figure 3-1 was used. In this model, Z_{sg} and Z_{vt} are transmission line networks, as in Figure 2-1. The asymmetric model of the lower airway is used to calculate Z_{sg}, unless otherwise noted. This model of the coupling between upper and lower airways was first proposed by Stevens [48].

3.1.1 *The glottis*

The glottal impedance Z_g was calculated using Equation 3.1 [56, 48], where μ = coefficient of viscosity, d = width of the glottal slit, l_g = length of the glottal opening, h = thickness of the glottis, U_g = volume flow through the glottis, and K = a constant close to unity. The values of these variables used in the model are given in Table 3.1,

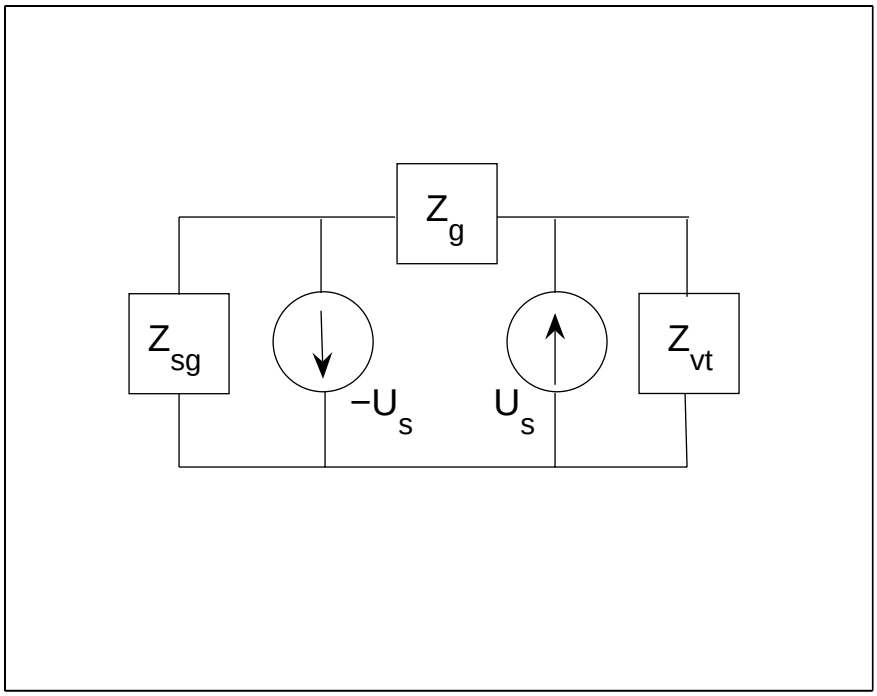


Figure 3-1: Circuit of the vocal tract, including subglottal impedance Z_{sg}, vocal tract impedance Z_{vt}, and two opposite-polarity volume velocity sources U_s straddling the glottal impedance Z_g.

Table 3.1: Parameters for calculating the glottal impedance Z_g

Parameter	Value
μ	$1.8 \cdot 10^{-1}$
h	$4mm$
l_g	$18mm$
d	variable
K	0.875

and are based on values found in the literature [56, 16, 48].

The model is limited by the assumption that speech production is accomplished by a linear source-filter system. DC flow is assumed to be zero at all times, the formation of vortices and/or sheet flow in the vocal tract is assumed not to occur, and source excitation of the filter is calculated independently of vocal fold vibration. The source spectrum was modeled at unity for all frequencies. The calculated vowel spectra are therefore in effect simply the impulse responses of the system for given configurations of the supraglottal vocal tract and larynx, with impulses occurring simultaneously on each side of the glottal impedance and with opposite polarity (see Appendix A for the MATLAB code implementation).

$$Z_g = \left[\frac{12\mu h}{(l_g d^3)} + K\frac{\rho U_g}{(l_g d)^2}\right] + j\omega\frac{\rho h}{l_g d} \tag{3.1}$$

3.1.2 *The vocal tract*

The vocal tract transmission line was modeled so that each T-network represented a uniform tube with rigid walls and length no greater than 1.5 cm. Acoustic losses due to air friction were incorporated in the same way that they were in the subglottal model (cf. Figure 2-1). Vocal tract area functions were taken from Story [51], speaker SM1.

3.1.3 *Properties of the model*

The model in Figure 3-1 has several properties which should be noted. First, in the case when the glottal impedance is infinite ($Z_g \rightarrow \infty$), the result is a volume velocity circulation in the vocal tract and a separate volume velocity circulation in the lower airway. The two volume velocity waveforms have identical shape and magnitude but opposite polarity. This result is a direct effect of implementing two sources straddling the glottal impedance rather than just one source to the right of it. Physiologically, this arrangement conforms to our expectations, at low frequencies, since the buildup of pressure below the glottis occurs in phase with the release of pressure above the glottis in real speech, and vice versa.

Second, for finite glottal impedances and for frequencies at which the lower airway impedance is infinite (at a pole), the left side of the circuit is open. Writing the node equation for the node just below the right-hand source yields $U_S + -U_S + -U_{vt} = 0$, in which the positive U_S refers to the right-hand source, the negative U_S refers to the left-hand source, and the U_{vt} refers to the volume velocity through the vocal tract. Since the three volume velocities must add to zero, the volume velocity through the vocal tract, U_{vt}, must also be zero. Likewise, for frequencies at which the lower airway impedance is zero (at a zero), the left side of the circuit is shorted and $U_{vt} = U_S$. Since losses are built into the model, however, the lower airway impedance ranges not between infinitely large and infinitely small, but between very large and very small. At frequencies near the poles of the lower airway impedance, the effect of the lower airway on the vowel spectrum is maximum (decreasing the amplitude of the vowel spectrum at those frequencies). At frequencies near the zeros of the lower airway impedance, the effect of the lower airway on the vowel spectrum is minimum.

The poles of the lower airway impedance therefore appear as zeros in the vowel spectrum. The converse is not true, however, since the poles of the vowel spectrum

are determined by the whole system (lower airway plus vocal tract) and not by the lower airway alone.

3.2 Results

3.2.1 *General results*

Figures 3-2 through 3-5 show the effects on vowel spectra from gradually decreasing the glottal impedance by manipulating the glottal width parameter d. Figure 3-6 shows the F1∼F2 chart for the eleven English vowels for which Story [51] provides area function data. Table 3.2 gives the data in numerical form.

For the case $Z_g \rightarrow \infty$, the spectra of the vowels and their formant frequencies are roughly what we expect. As Z_g decreases, the formants generally increase in frequency while their amplitudes decrease. The decrease in amplitude cannot be attributed entirely to the introduction of zeros into the spectrum, however. The vocal tract walls were modeled as being rigid. Increased coupling to the non-rigid lower airway model should therefore be expected to result in a global decrease of amplitude because of the larger resistive component. The frequencies of the formants will also be affected by this property of the model. Figure 3-7 shows the spectra of the vowels [i], [a], and [u] in the case of the lower airway model with rigid walls, and Table 3.3 gives the frequencies and amplitudes of the formants.

As noted in Chapter 2, the symmetric model of the lower airway results in an input impedance with more prominent poles and zeros than the asymmetric model. The results are similar in the vowel spectra. Figure 3-8 shows the spectra for the vowels [i], [a], and [u] when the lower airway is symmetric.

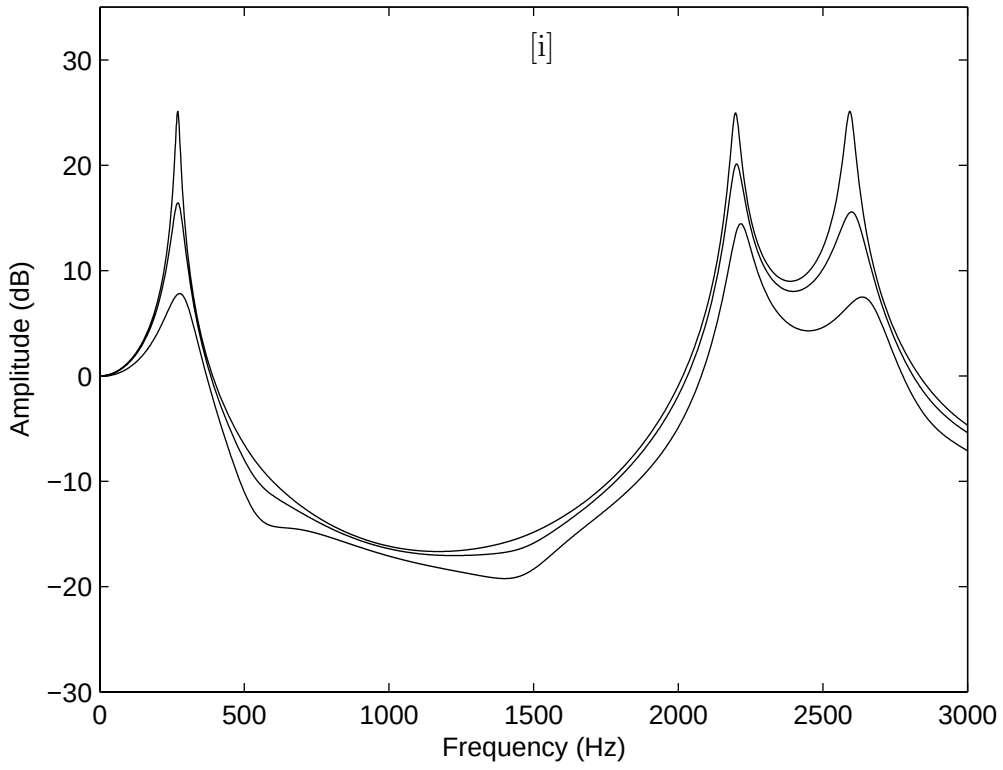


Figure 3-2: Three spectra of [i] as the glottal impedance Z_g decreases. Spectra with lower amplitudes indicate a smaller Z_g and hence more coupling between the lower airway and the vocal tract. The three Z_g conditions correspond to glottal widths of $d = 0.6mm$, $1.2mm$, and $1.8mm$.

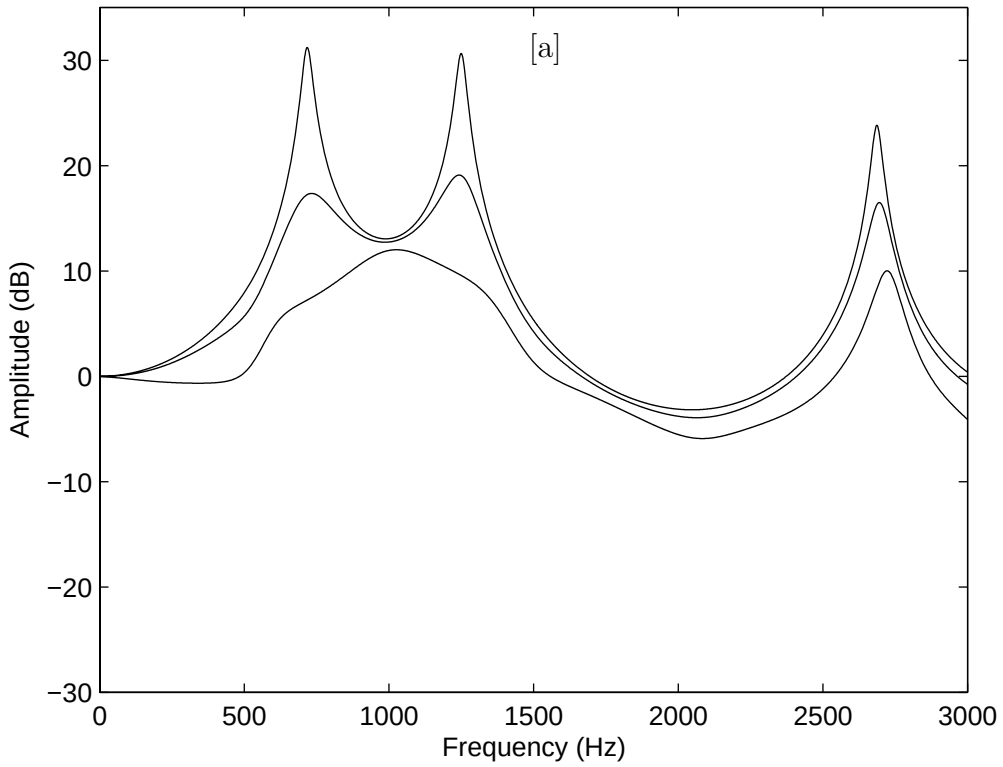


Figure 3-3: Three spectra of [a] as the glottal impedance Z_g decreases. The three Z_g conditions correspond to glottal widths of $d = 0.6mm$, $1.2mm$, and $1.8mm$.

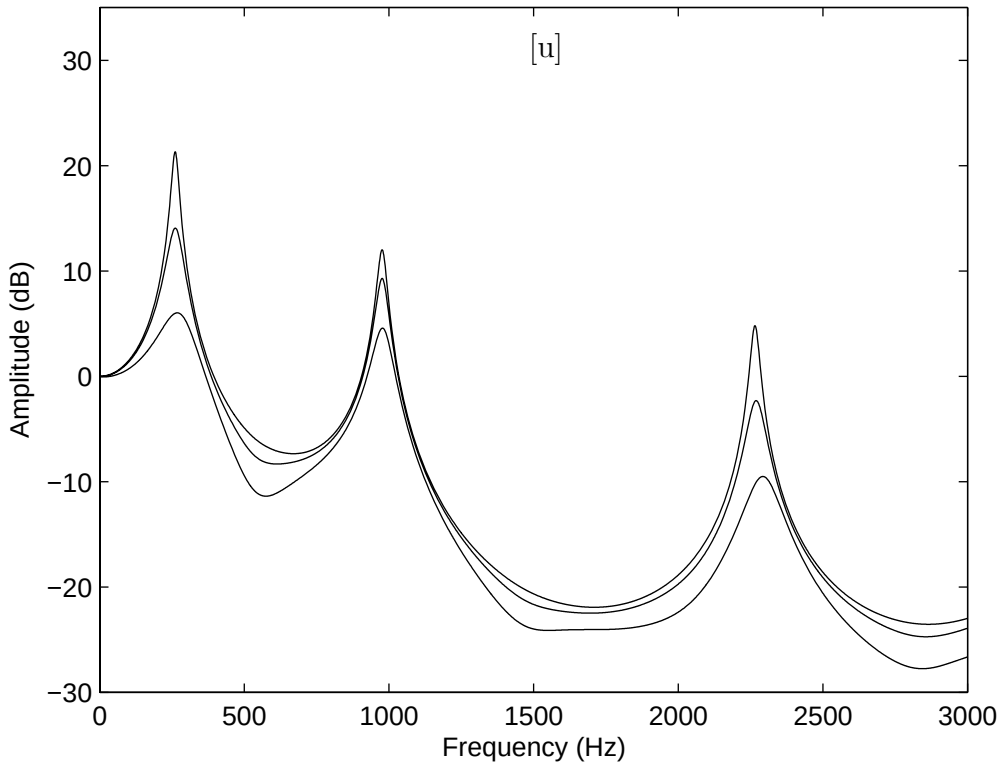


Figure 3-4: Three spectra of [u] as the glottal impedance Z_g decreases. The three Z_g conditions correspond to glottal widths of $d = 0.6mm$, $1.2mm$, and $1.8mm$.

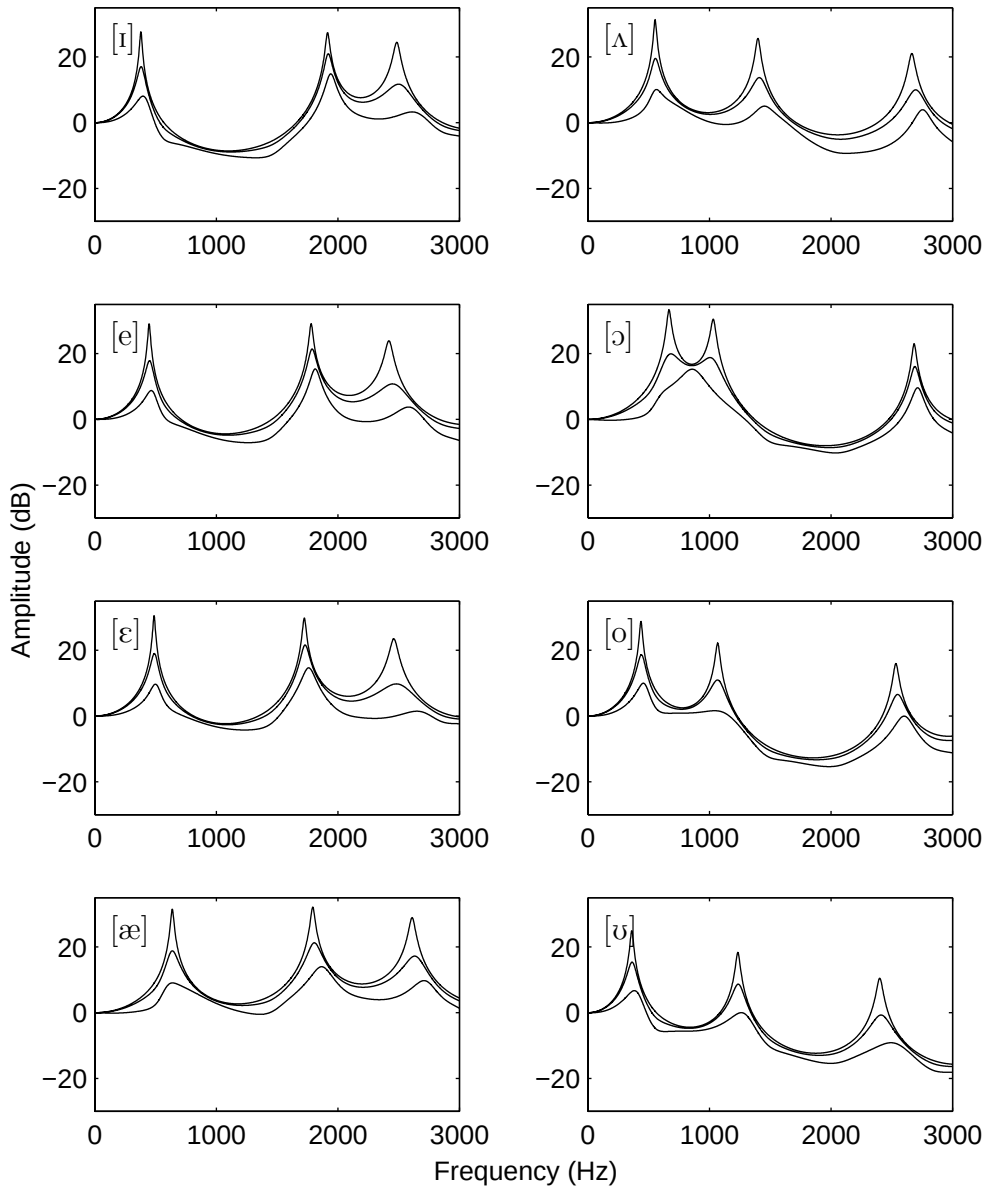


Figure 3-5: Three spectra of the vowels [ɪ], [e], [ɛ], [æ], [ʌ], [ɔ], and [ʊ] as the glottal impedance Z_g decreases. The three Z_g conditions correspond to glottal widths of $d = 0.6mm$, $1.2mm$, and $1.8mm$.

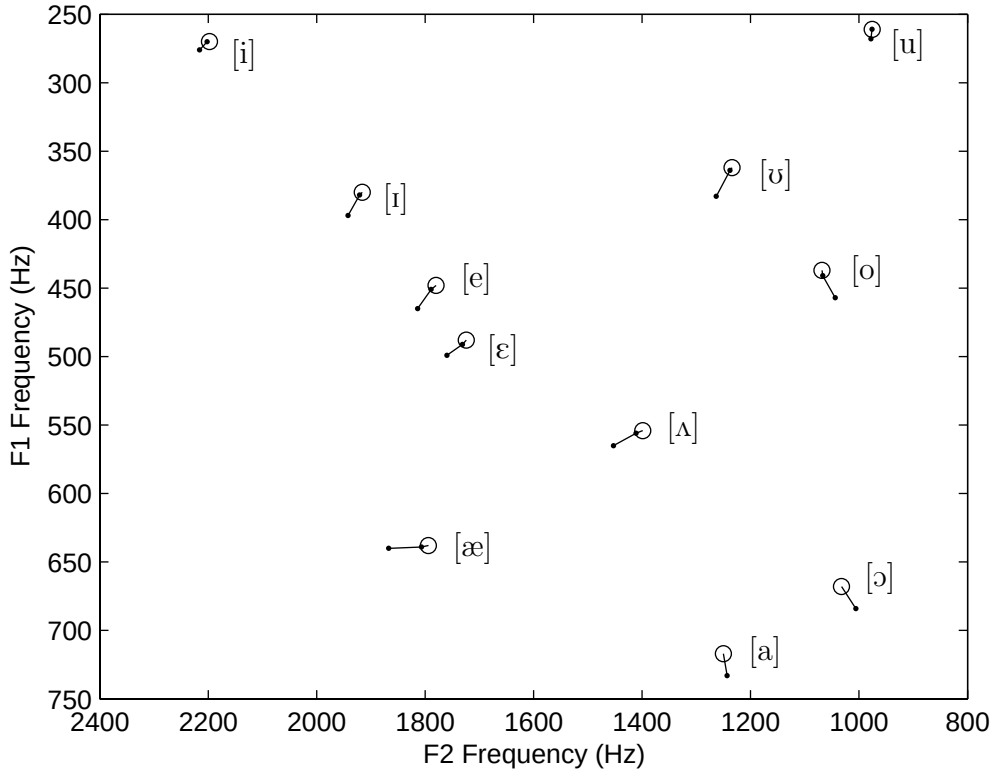


Figure 3-6: F1∼F2 vowel chart. The open circles indicate the frequencies of F1 and F2 when the glottal width is $d = 0.6mm$. As d increases to $1.8mm$, the formants move to frequencies indicated by the dots. (There are 2 dots per vowel, except for [a] and [ɔ]; numerical data is given in Table 3.2.)

Table 3.2: Frequencies and amplitudes of the first three formants for all eleven vowels and three values of the glottal width d.

		F1	F2	F3	A1	A2	A3
	[i]	270	2198	2593	25.1084	24.9658	25.1162
	[ɪ]	380	1916	2485	27.6502	27.4016	24.4612
	[e]	448	1780	2420	29.0812	29.0590	23.9180
	[ɛ]	488	1724	2461	30.4521	29.8406	23.5002
	[æ]	638	1794	2611	31.4887	32.1031	28.9394
$d = 0.6mm$	[ʌ]	554	1399	2665	31.3566	25.6713	21.0515
	[a]	717	1250	2687	31.2174	30.6535	23.8308
	[ɔ]	668	1032	2684	33.2823	30.4936	22.9303
	[o]	437	1068	2534	28.8017	22.1757	15.9468
	[ʊ]	362	1234	2401	24.9650	18.3023	10.5012
	[u]	261	976	2265	21.3236	12.0106	4.8097
	[i]	270	2202	2599	16.4258	20.1085	15.5698
	[ɪ]	382	1921	2500	17.0738	20.9886	11.7205
	[e]	451	1789	2451	17.8588	21.4271	10.7947
	[ɛ]	491	1731	2482	19.0239	21.6282	9.7720
	[æ]	639	1807	2633	18.8200	21.2757	17.2286
$d = 1.2mm$	[ʌ]	556	1411	2698	19.5682	13.7149	9.9927
	[a]	733	1243	2695	17.3580	19.1022	16.4976
	[ɔ]	684	1006	2691	19.9402	18.8057	16.0684
	[o]	441	1067	2549	18.6717	10.9705	6.5687
	[ʊ]	364	1238	2412	15.4046	8.6891	-0.7110
	[u]	261	976	2269	14.0660	9.2947	-2.3130
	[i]	276	2216	2636	7.8114	14.4417	7.4843
	[ɪ]	397	1942	2612	8.0834	14.8644	3.2786
	[e]	465	1814	2583	8.7746	15.3521	3.7487
	[ɛ]	499	1760	2651	9.6991	14.6702	1.4616
	[æ]	640	1867	2710	9.0569	13.9925	9.7630
$d = 1.8mm$	[ʌ]	565	1453	2755	10.1085	5.0394	3.9431
	[a]			2722			10.0095
	[ɔ]			2713			9.6187
	[o]	457	1044	2603	9.9912	1.6118	-0.0335
	[ʊ]	383	1263	2495	6.7115	0.0059	-9.1455
	[u]	268	978	2292	6.0218	4.5742	-9.5138

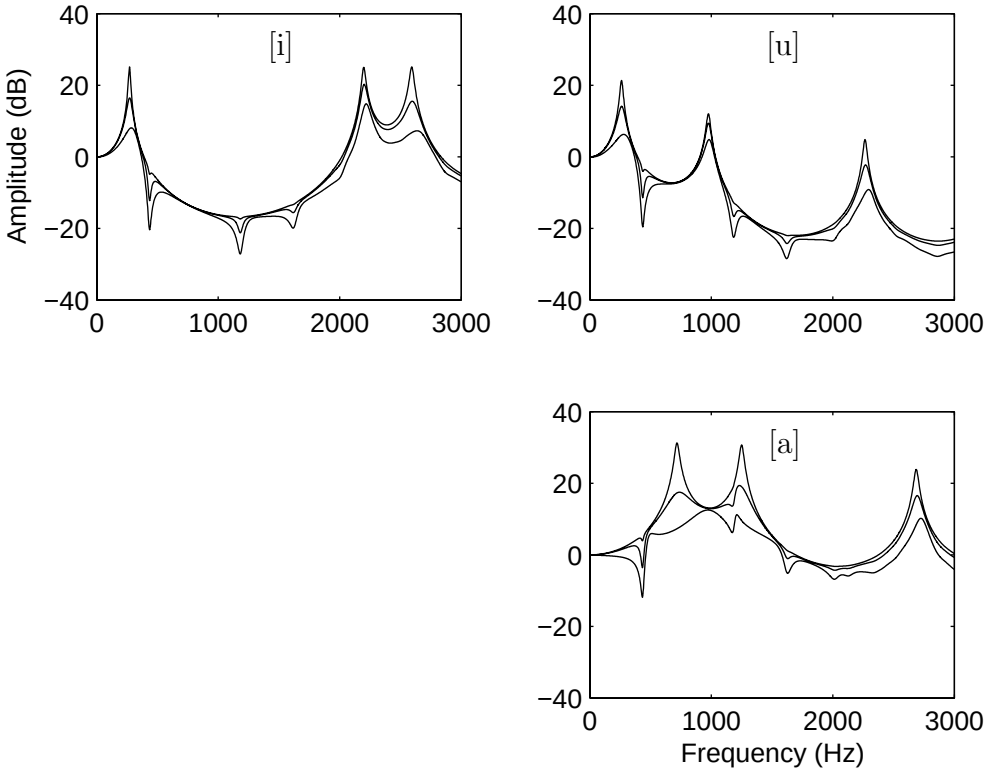


Figure 3-7: Three spectra for the vowels [i], [u], and [a], in the case that the lower airway is modeled with rigid walls. Table 3.3 gives the numerical values for the frequencies and amplitudes of the first three formants.

Table 3.3: Frequencies and amplitudes of the first three formants for the vowels [i], [u], [a] and three values of the glottal width d. For these data, the lower airway was modeled with rigid walls.

		F1	F2	F3	A1	A2	A3
	[i]	270	2198	2593	25.1256	24.9913	25.1150
$d = 0.6mm$	[u]	261	976	2265	21.3409	12.0319	4.8276
	[a]	717	1250	2687	31.2499	30.6486	23.8513
	[i]	271	2202	2598	16.4830	20.2341	15.5425
$d = 1.2mm$	[u]	262	977	2270	14.1283	9.4039	-2.2324
	[a]	738	1232	2695	17.4925	19.3664	16.5602
	[i]	286	2218	2638	8.1014	14.8393	7.2808
$d = 1.8mm$	[u]	279	981	2295	6.2756	4.8408	-9.1324
	[a]	972	1210	2725	12.5155	11.2218	10.2023

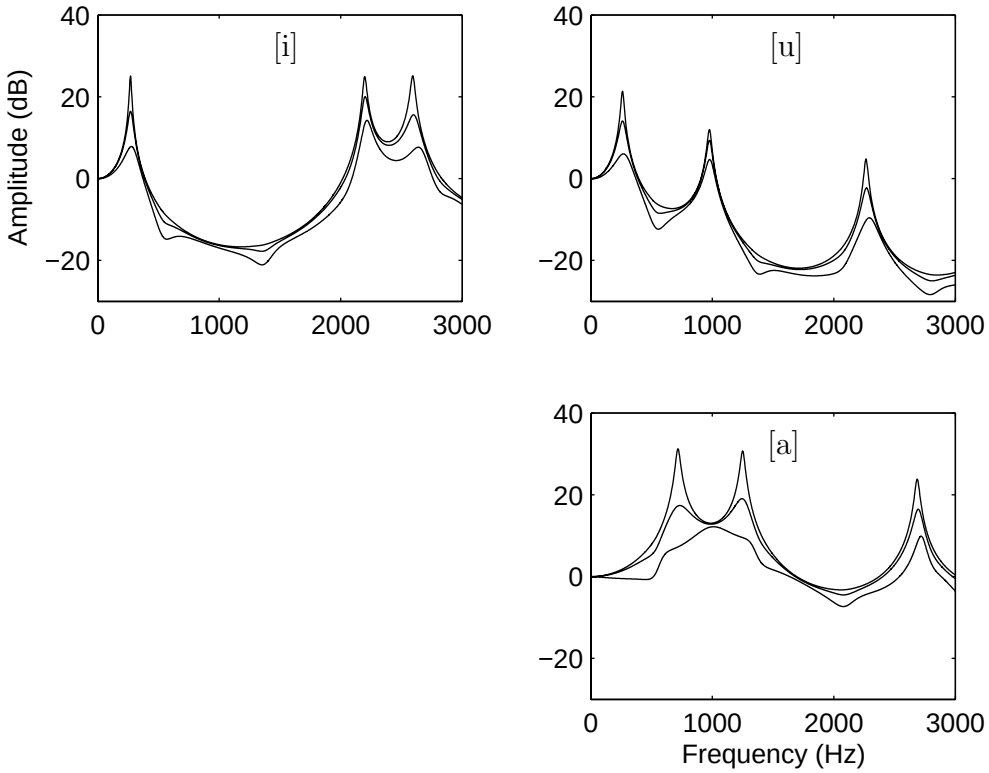


Figure 3-8: Three spectra for the vowels [i], [u], and [a], in the case that the lower airway is symmetric.

3.2.2 *Relation of formants to lower airway resonances*

Comparison of Tables 2.12 and 3.2 shows the relative distribution of vowel formants with respect to the lower airway resonances. F1 for the [+low] vowels [æ], [a], [ʌ], and [ɔ] lies between the first and second lower airway resonances, while for [-low] vowels it lies below the first lower airway resonance. F2 for [-back] vowels lies between the second and third lower airway resonances, and for [+back] vowels F2 lies between the first and second lower airway resonances. F2 for [i] lies above the third lower airway resonance.

It should be noted that these data can be suggestive only, not conclusive. The lower airway model was based on an anatomical study of adult male lungs by Horsfield et al [24], and the vocal tract model was based on an MRI study of a single adult male by Story [51]. Each of these parts of the total model has its own advantages and disadvantages, and the combination of the two parts likewise has advantages and disadvantages. In Horsfield's [24] study, data were collected from dead lungs which may have undergone morphological changes between the time of death and the time of examination. Indeed, this has been suggested as a reason for the discrepancies between van den Berg's [55] and Ishizaka's [26] data (van den Berg [55] used cadavers, whereas Ishizaka et al [26] used living tracheotemized patients). Story's [51] data on the vocal tract are limited by the low time resolution of MRI. In his study, subjects produced prolonged utterances of each of the vowels so that the the recorded images would not blur too significantly as the tongue and other articulators move. Further complicating the matter is the fact that in English [e], [o], and [u] are heavily diphthongized. When these vowels are prolonged, it is not known how the nucleus and the offglide of the diphthongs are merged, or whether one simply drops out. In addition, subjects lie down when inside the magnet, and must therefore speak in an unusual posture. It is not clear exactly how vowel production under these conditions differs

from normal vowel production. Finally, there is no way of knowing how closely the lower airway of Story's subject matches the model of the lower airway employed here. It will be necessary, therefore, to take these results as suggestive only.

3.3 Further remarks

The sensitivity of F1 and F2 to lower airway coupling in the back low vowels [a] and [ɔ] requires some explanation. In these vowels, the pharyngeal cavity is highly constricted, so that the impedance of the pharynx is closer to the impedance of the glottis. The pharynx and the lower airway are therefore acoustically more tightly coupled in these vowels than in other vowels. Because the pharynx opens gradually into the oral cavity, the coupling between the lower airway and the whole vocal tract is enhanced. It is for this reason that the same glottal width d leads to greater effects on the spectra of the low back vowels [a] and [ɔ] than it does on the spectra of the other vowels. This raises several questions. First, it may be that [-ATR] (or [+RTR]) vowels show a greater degree of coupling to the lower airway than [+ATR] (or [-RTR]) vowels do, leading to more prominent pole-zero pairs. Second, it may be that different voice qualities also exhibit varying degrees of lower airway coupling. For instance, in pharyngealized or creaky voice, the pharynx and larynx are constricted to form a relatively narrow tube above the glottis [12], which may affect the coupling to the lower airway in addition to source-filter interactions. In whispered speech, formants are generally seen to rise, especially F1 and especially in the low back vowels [31, 32]. This could be due to the increase in coupling as the glottal width d increases.

Finally, the region near F1 and F2 in the vowel [a] may be of use to voice pathologists seeking to evaluate voice function in patients, since this region appears to be the most sensitive to changes in glottal impedance. A set of parameters could be devised which might correlate either with the parameters H1-H2 and A1-A3 [50] or

with physiological measures of vocal disfunction in pathologies that interfere with glottal closure.

3.3.1 *Major conclusions of Chapter 3*

The frequencies of the vowel formants tend to increase as the glottis is abducted. This mirrors the formant data from whispered vowels [31, 32], which are produced with a more abducted glottis. Furthermore, the region around F1 and F2 in low back vowels is very sensitive to changes in glottal abduction, and can be dominated by the pole of the first lower airway resonance. This sensitivity could potentially be used to evaluate voice pathologies that interfere with glottal closure during vocal fold vibration.

Front and back vowels seem to be separated in the F2 plane by the second lower airway resonance. Low and non-low vowels seem similarly to be separated in the F1 plane by the first lower airway resonance. The interaction of the first and third lower airway resonances with the second formant in [+ATR] and [-ATR] vowels is less clear. In Chapter 4, the relation between F2 and the second and third lower airway resonances in in vowels and consonants will be explored in an acoustic study.

Chapter 4

Measurements on Consonant and Vowel Spectra

4.1 Method

Speech from one adult male speaker was analyzed. The speech samples were obtained as part of a free database collected at the Language Technologies Institute at Carnegie Mellon University (the ARCTIC database [38]), which includes over 1000 sentences of read speech.

405 CV transitions were measured. In all cases the vowel of the CV was a full vowel, usually in a stressed syllable, and the following segment was a non-sonorant consonant. (There were some CVs which were followed by [h] or a vowel, but only if there was a clear steady-state portion of the V. In some cases where a steady-state portion of the vowel could not be determined due to excessive coarticulation, the CV was left out of the analysis.) No CVs were measured which were followed by a liquid or nasal consonant, so that the additional poles and zeros from those consonants would not be present. The tokens from the ARCTIC database that were analyzed

Table 4.1: Summary of the number and type of CV tokens analyzed from the ARCTIC database. All six stops were analyzed in the context of 9 following vowels.

C/V	[i]	[ɪ]	[ɛ]	[æ]	[ʌ]	[a]	[ɔ]	[ʊ]	[u]	Total
[b]	32	21	7	20	26	4	7	3	0	120
[p]	7	6	2	8	3	2	2	3	0	33
[d]	7	40	11	2	5	13	0	0	9	87
[t]	6	0	2	8	1	8	0	8	27	60
[g]	1	9	5	2	0	10	0	13	2	42
[k]	5	1	0	22	3	10	2	20	0	63

are given in Appendix B, and a summary of the number of tokens per CV type is given in Table 4.1.

Spectral measurements were made at two locations in the CVs: 1) $F2_{onset}$, at the earliest time in which F2 could be measured, using a 256 ms Hamming window, and 2) $F2_{vowel}$, in the steady state or middle part of the vowel, using a 1024 ms Hamming window. The sampling rate was 32000 Hz. Determination of the steady state or middle part of the vowel depended on the consonantal context both preceding and following the vowel, but was generally straight-forward. Measurements of the formant track discontinuity were also made during the F2 transitions of a subset of the tokens. These measurements include 1) $F2_{high}$, the F2 on the high frequency side of the discontinuity, and 2) $F2_{low}$, the F2 on the low frequency side of the discontinuity.

4.2 The relationship between lower airway resonances and vowel spectra

A histogram of the $F2_{vowel}$ measurements is plotted in Figure 4-1. There is a small number of $F2_{vowel}$ measurements falling in the regions near the lower airway resonances, whereas many $F2_{vowel}$ measurements fall between the lower airway resonances. [-back] vowels have $F2_{vowel}$ higher than the second lower airway resonance,

while [+back] vowels have $F2_{vowel}$ lower than the second lower airway resonance. Furthermore, $F2_{vowel}$ for the [+ATR] front vowels is higher than the third lower airway resonance, while it is lower than the third lower airway resonance for [-ATR] front vowels. The [+ATR] and [-ATR] back vowels are not so easily distinguished on the basis of lower airway resonances. For both classes of back vowels, $F2_{vowel}$ falls between the first and second lower airway resonances. It is worth noting, however, that the $F2_{vowel}$ in [+ATR] back vowels falls closer to (or lower than) a pole that appears near 1000 Hz. It is still unclear what gives rise to this pole, but it is at about the same frequency as the first lower airway pole in vowels such as [a] and [ɔ], as was shown in Chapter 3. If this pole is a result of the coupling between the upper and lower airways, it may be possible to define [±ATR] in back vowels according to the location of $F2_{vowel}$ with respect to this pole.

The frequencies of $F2_{high}$ and $F2_{low}$ during CV transitions are given in Table 4.2. The second lower airway resonance of this speaker can be estimated by the frequency midway between the mean $F2_{high}$ and $F2_{low}$ measurements. This estimated frequency ($\frac{1632Hz-1341Hz}{2} = 1486Hz$) is only 16 Hz higher than the frequency calculated in Chapter 2, and therefore the modeled frequencies of the second and third lower airway resonances will be used for comparisons with the speaker's second formant.

Figure 4-2 gives some examples of the kinds of discontinuity that are represented in Table 4.2.

4.3 The relationship between lower airway resonances and consonant-vowel transitions

Figure 4-3 plots $F2_{onset}$ vs. $F2_{vowel}$ for the speaker. Such plots are used to determine locus equations for consonant places of articulation [11, 39, 52, 18]. Locus equations

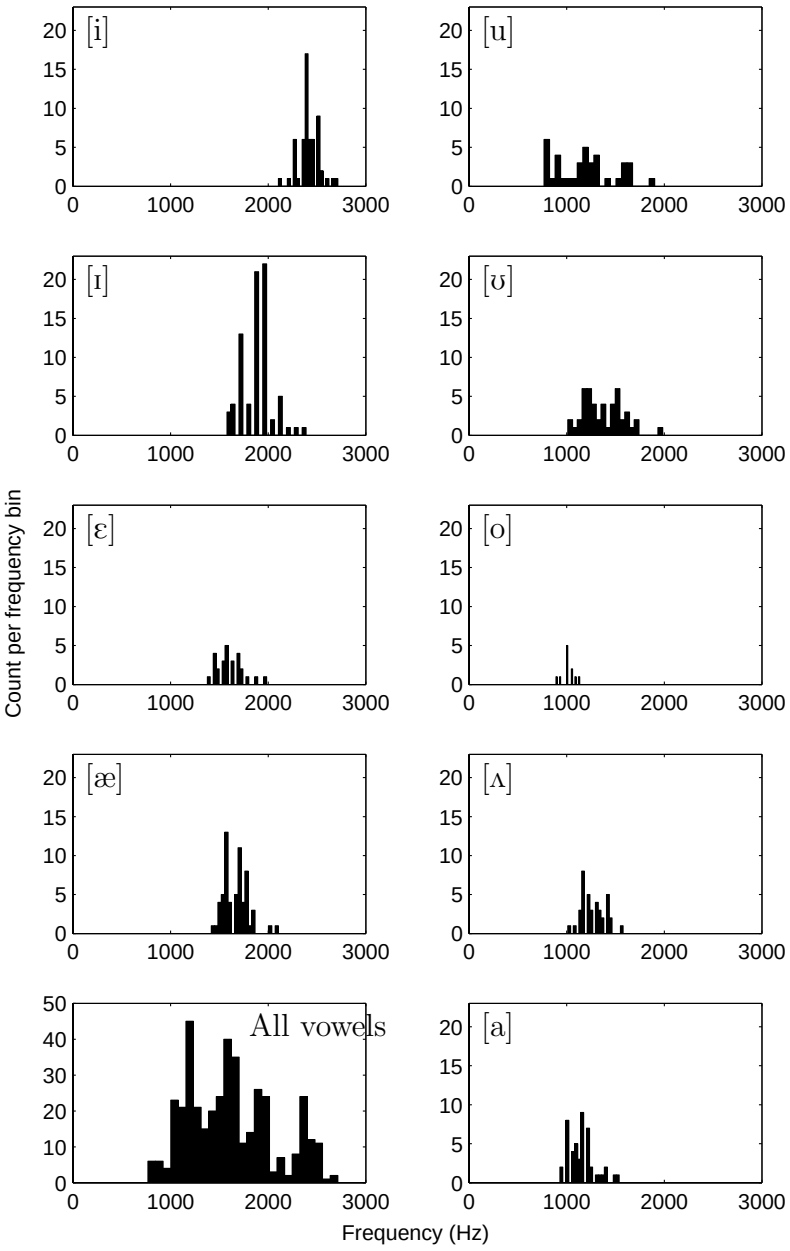


Figure 4-1: Histograms of $F2_{vowel}$ for each vowel, and for all vowels together (bottom left panel). Front vowels are given in the left column and back vowels are given in the right column.

Table 4.2: Frequencies of $F2_{high}$ and $F2_{low}$ for a handful of tokens in which a discontinuity in F2 was very clear. Examples of such discontinuities are given in Figure 4-2. Such clear discontinuities occur most frequently in transitions from a [d] or [t] to a back vowel. The token label indicates the name of the file in the ARCTIC database where the measurements were made.

	$F2_{high}$	$F2_{low}$	Label
[dʌ]	1620	1377	a0072
	1498	1377	a0261
	1498	1377	a0524
[da]	1498	1255	a0197
	1620	1255	b0239
	1539	1377	b0466
[du]	1498	1296	a0368
	1944	1255	b0462
	1863	1377	b0226
[tʊ]	1579	1255	a0494
	1620	1377	a0497
	1498	1377	b0476
[bi]	1651	1395	a0280
	1924	1417	a0508
Mean	1632	1341	

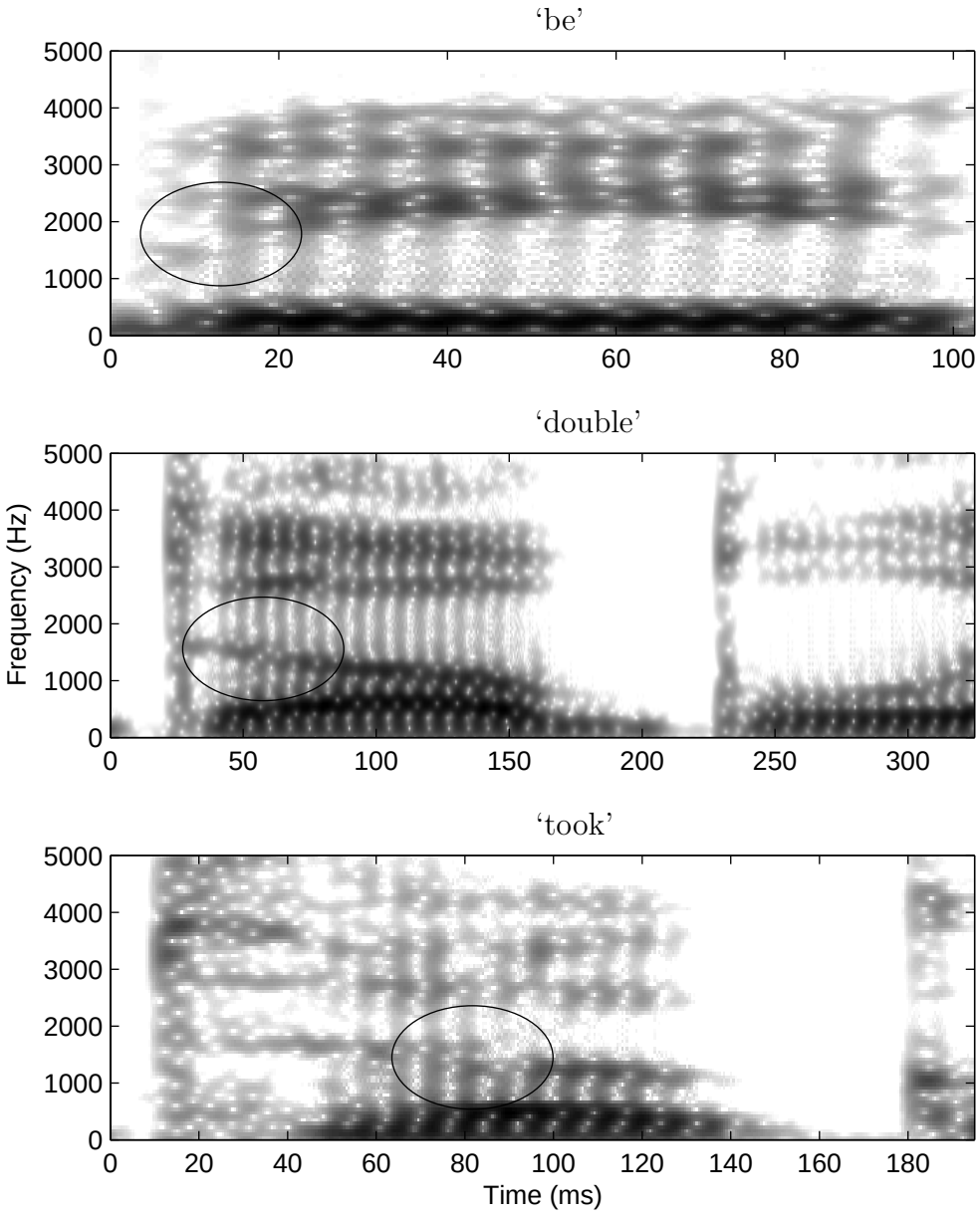


Figure 4-2: Examples of clear discontinuities as F2 cross the second lower airway resonance. Top panel: F2 rises abruptly at the beginning of the vowel in the word 'be'. Middle panel: F2 falls abruptly at the beginning of the vowel in the word 'double'. Bottom panel: F2 falls abruptly during the vowel in the word 'took'.

Table 4.3: Slopes and y-intercepts for the locus equations describing the scatter plots in Figure 4-3. Locus equations were calculated for each phoneme both independent of vowel context, and dependent on whether the following vowel was [-back] or [+back].

Phoneme	slope	y-intercept (Hz)
[b]	0.49	595
[p]	0.43	885
[d]	0.30	1395
[t]	0.20	1665
[g]	1.02	288
[k]	0.95	454
[b] before a front vowel	0.30	1071
[b] before a back vowel	0.38	674
[p] before a front vowel	0.00	1823
[p] before a back vowel	-1.70	3093
[d] before a front vowel	0.40	1301
[d] before a back vowel	0.10	1614
[t] before a front vowel	0.30	1445
[t] before a back vowel	0.00	1872
[g] before a front vowel	0.20	1885
[g] before a back vowel	0.30	1101
[k] before a front vowel	0.40	1690
[k] before a back vowel	0.66	689

are linear regression equations for the $F2_{onset}$ ~$F2_{vowel}$ scatter plots by consonant place of articulation, and are characterized by the y-intercepts and slopes of these lines. In general, the alveolar stops [d,t] have the shallowest slope, while the velars [g,k] and labials [b,p] have steeper slopes. The velars and the labials are distinguished from each other by the y-intercept, which is higher for the labials than for the velars. The same overall pattern is seen in Figure 4-3. (See Table 4.3 for the slopes and y-intercepts for the speaker.) There is, however, some further structure to the scatter plots, some of which has been described by other researchers and some of which has not.

The structure of the scatter plot for velars is more complicated, as is well known.

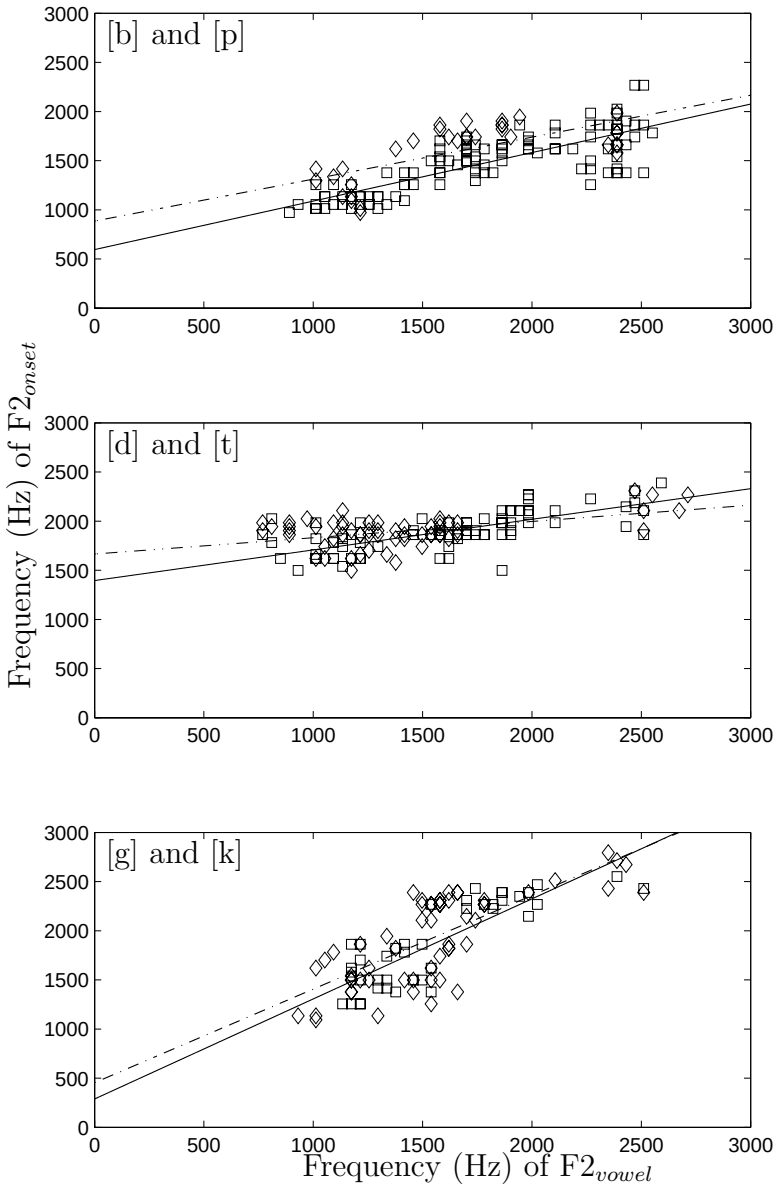


Figure 4-3: Scatter plots of $F2_{onset}$ vs. $F2_{vowel}$ and regression lines (locus equations). Solid lines are the regression lines for the voiced stops (squares), dashed lines are the regression lines for the voiceless stops (diamonds). Top panel: Labial stops in all vowel contexts. Middle panel: Alveolar stops in all vowel contexts. Bottom panel: Velar stops in all vowel contexts.

Before front vowels the velars are fronted, while they are backed before back vowels. This gives rise to a bimodal distribution of burst and formant transition properties. Sussman et al. [52] takes account of this and analyzes the locus equations for fronted velars and backed velars separately.

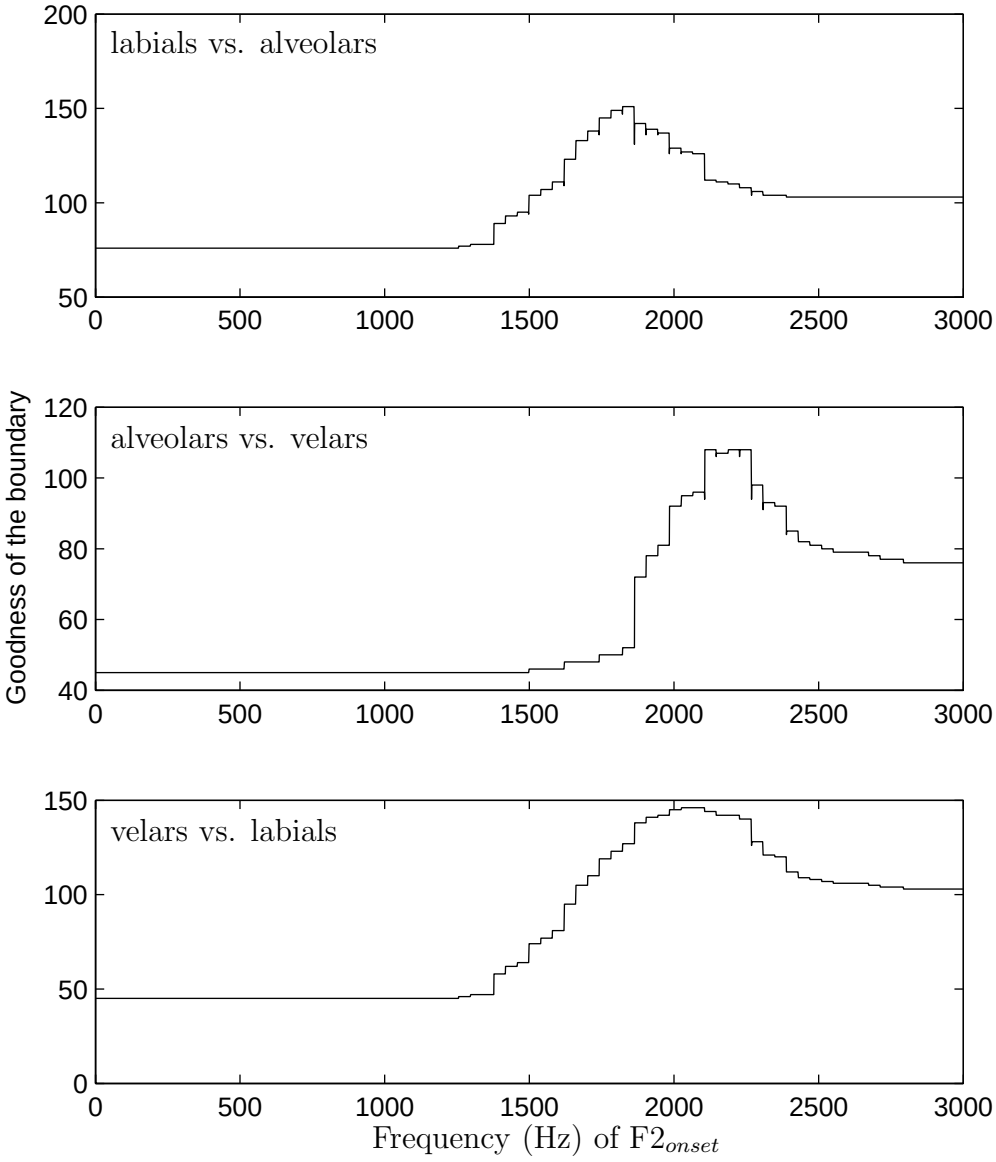
The slope of the locus equation for velars in a back vowel context is small, and it is even smaller for velars in a front vowel context. In back vowel contexts, $F2_{onset}$ for velars is generally between the second and third lower airway resonance, while in front vowel contexts it is above the third lower airway resonance. Similarly, $F2_{onset}$ for alveolars is generally between the second and third lower airway resonances. $F2_{onset}$ for labials generally lies below the second lower airway resonance in back vowel contexts, and between the second and third lower airway resonances in front vowel contexts.

4.3.1 *Boundaries between consonant places of articulation*

Figures 4-4 and 4-5 plot the results of a simple statistical analysis of the scatter plots for front and back vowel contexts. For each hypothesized $F2_{onset}$ boundary frequency (ranging from 0 to 3000 Hz in 1 Hz increments), the number of [αPlace] data points falling below that boundary plus the number of [βPlace] data points falling above that boundary is taken to be a measure of the goodness of the boundary. Peaks in the curve therefore represent optimal frequencies for the boundary.

For instance, if labials are expected to have lower $F2_{onset}$ than alveolars, then for a hypothesized boundary frequency f the goodness of this boundary is calculated as the number of labials with $F2_{onset}$ less than f plus the number of alveolars with $F2_{onset}$ greater than f. For the comparison of labials and velars, labials are again assumed to have a lower $F2_{onset}$ than velars. Alveolars are assumed to have a lower $F2_{onset}$ than velars.

The boundaries picked out by this simple statistical analysis coincide with the


labials vs. alveolars
alveolars vs. velars
velars vs. labials
Goodness of the boundary
Frequency (Hz) of F2$_{onset}$

lower airway resonances. Specifically, the boundary between labials and alveolars is around the second lower airway resonance for back vowel contexts. For front vowel contexts the boundary between labials and alveolars is shifted upward. The boundary between alveolars and velars is around the second lower airway resonance for back vowel contexts, and the third lower airway resonance for front vowel contexts. The boundary between labials and velars is around the second lower airway resonance in back vowel contexts, and the third lower airway resonance in front vowel contexts.

Figure 4-4 *(facing page)*: As the frequency of a boundary between two consonant categories (e.g. [b] and [d]) varies from 0 to 3000 Hz, the number of consonants from the first category (e.g. [b]) with $F2_{onset}$ below the boundary plus the number of consonants from the second category (e.g. [d]) with $F2_{onset}$ above the boundary is plotted along the y-axis as a measure of the goodness of the boundary between the two consonant categories. Top panel: Labials [b] and [p] vs. alveolars [d] and [t] in a front vowel context. The labials were assumed to have $F2_{onset}$ below the boundary and the alveolars were assumed to have $F2_{onset}$ above the boundary. The peak indicates the boundary which produces the best separation between labials and alveolars in a front vowel context. Middle panel: Alveolars [d] and [t] vs. velars [g] and [k] in a front vowel context. The alveolars were assumed to have $F2_{onset}$ below the boundary and the velars were assumed to have $F2_{onset}$ above the boundary. Bottom panel: Velars [g] and [k] vs. labials [b] and [p] in a front vowel context. The labials were assumed to have $F2_{onset}$ below the boundary and the velars were assumed to have $F2_{onset}$ above the boundary. The peaks in the middle and bottom panels occur near the third lower airway resonance. The peak in the top panel occurs between the second and third lower airway resonances, and could reflect either the second lower airway pole, or errors in the measurements of $F2_{onset}$ in labials and alveolars, or the lack of a lower airway resonance-defined boundary between labials and alveolars in front vowel contexts. The boundary would otherwise be expected to fall in the vicinity of the second lower airway resonance.

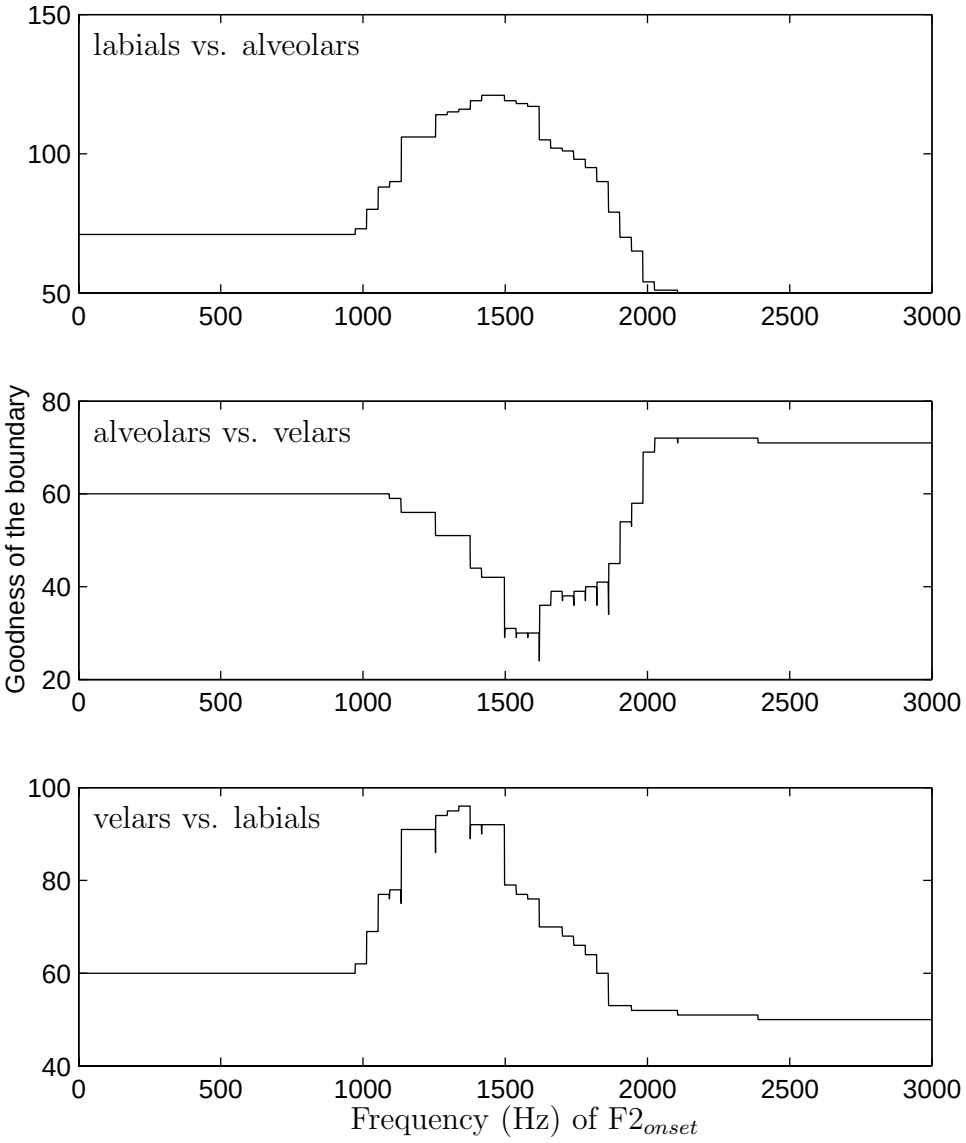


Figure 4-5: Same as in Figure 4-4, except the boundaries are for stop consonants in a back vowel context. The peaks in the top and bottom panels occur near the second lower airway resonance. In the middle panel there is a valley near the second lower airway resonance. There is no peak in the middle panel because the assumption about the relative frequencies of $F2_{onset}$ for alveolars and velars must be reversed. That is, in back vowel contexts $F2_{onset}$ in velars is below the boundary and $F2_{onset}$ in alveolars is above the boundary.

4.3.2 *A functional perspective on consonant-vowel transitions*

Not only do $F2_{vowel}$ and $F2_{onset}$ measures have a strong tendency to fall between the lower airway resonances (see Figures 4-1, and 4-6 through 4-8), but corresponding pairs of $F2_{vowel}$ and $F2_{onset}$ generally lie either well within a frequency band defined by consecutive lower airway resonances, or on opposite sides of a lower airway resonance. Figures 4-9 through 4-11 plot $F2_{onset}$ measures as 'x' and $F2_{vowel}$ measures as 'o'. Corresponding $F2_{onset}$ and $F2_{vowel}$ measures occur at the same height along the y-axis.

For velars before a front vowel (Figure 4-11, left column), $F2_{onset}$ is above the third lower airway resonance. $F2_{vowel}$ is below this resonance unless the vowel is [+ATR]. For alveolars and labials before a front vowel, both $F2_{onset}$ and $F2_{vowel}$ are between the second and third lower airway resonances. $F2_{vowel}$ is either above the third lower airway resonance (for [i]) or in the vicinity of the $F2_{onset}$.

For velars and alveolars before a back vowel $F2_{onset}$ is above the second lower airway resonance while $F2_{vowel}$ is below it. For labials before a back vowel both $F2_{onset}$ and $F2_{vowel}$ lie near each other and remain below the second lower airway resonance.

The following generalizations can be made, and may be of use in automatic speech recognition:

- 1) If F2 remains below the second lower airway resonance in a CV transition, the consonant must be labial and the vowel must be [+back].

- 2) If F2 crosses from above the second lower airway resonance to below it, the consonant must be either alveolar or velar, and the vowel must be [+back]. The distinction between alveolar and velar place may be made by reference to F3

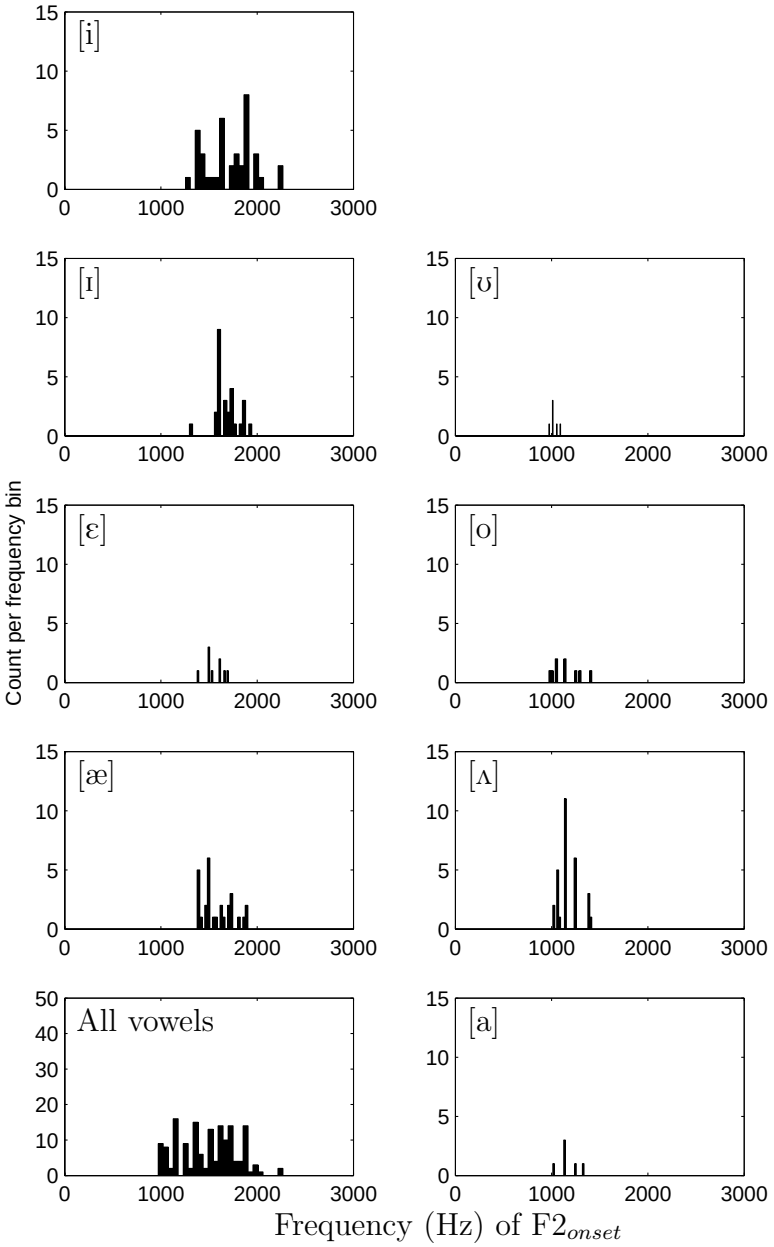


Figure 4-6: Histograms of F2$_{onset}$ for labials in each vowel context, and in all vowel contexts together (bottom left panel). Front vowels are given in the left column and back vowels are given in the right column.

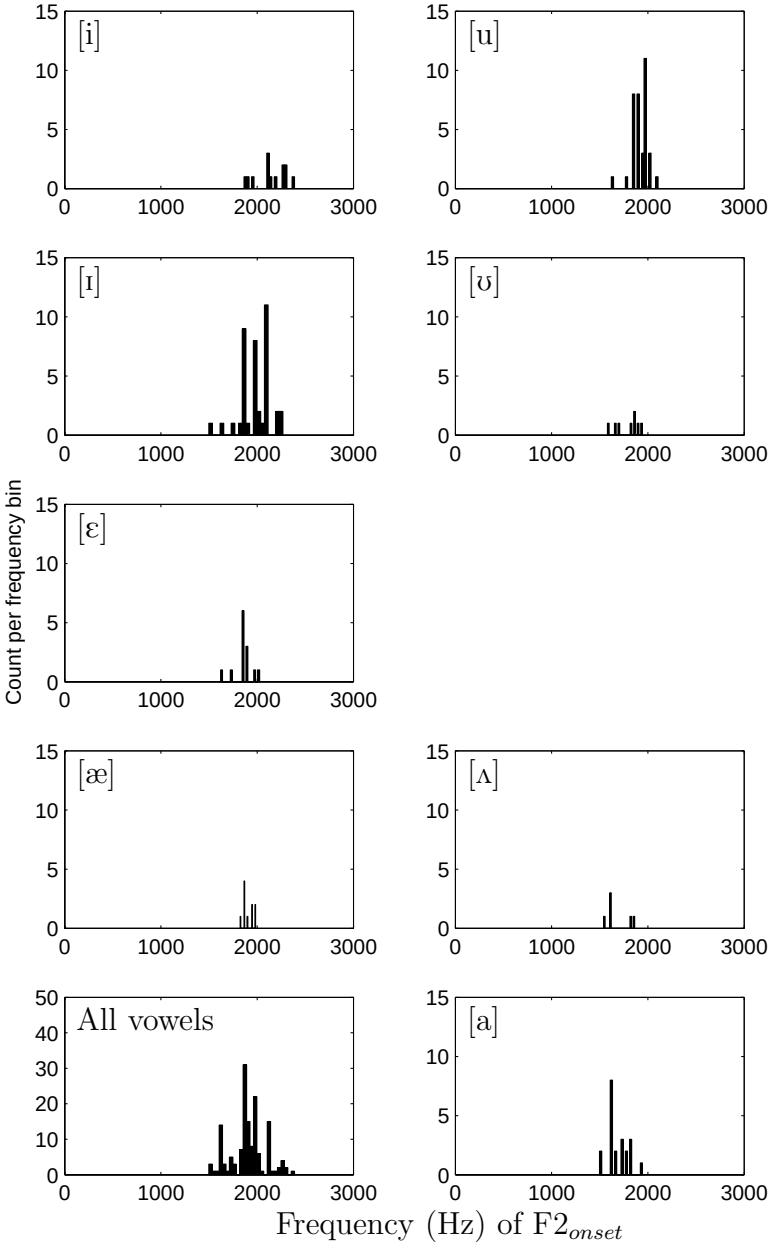


Figure 4-7: Same as in Figure 4-6 except for alveolars.

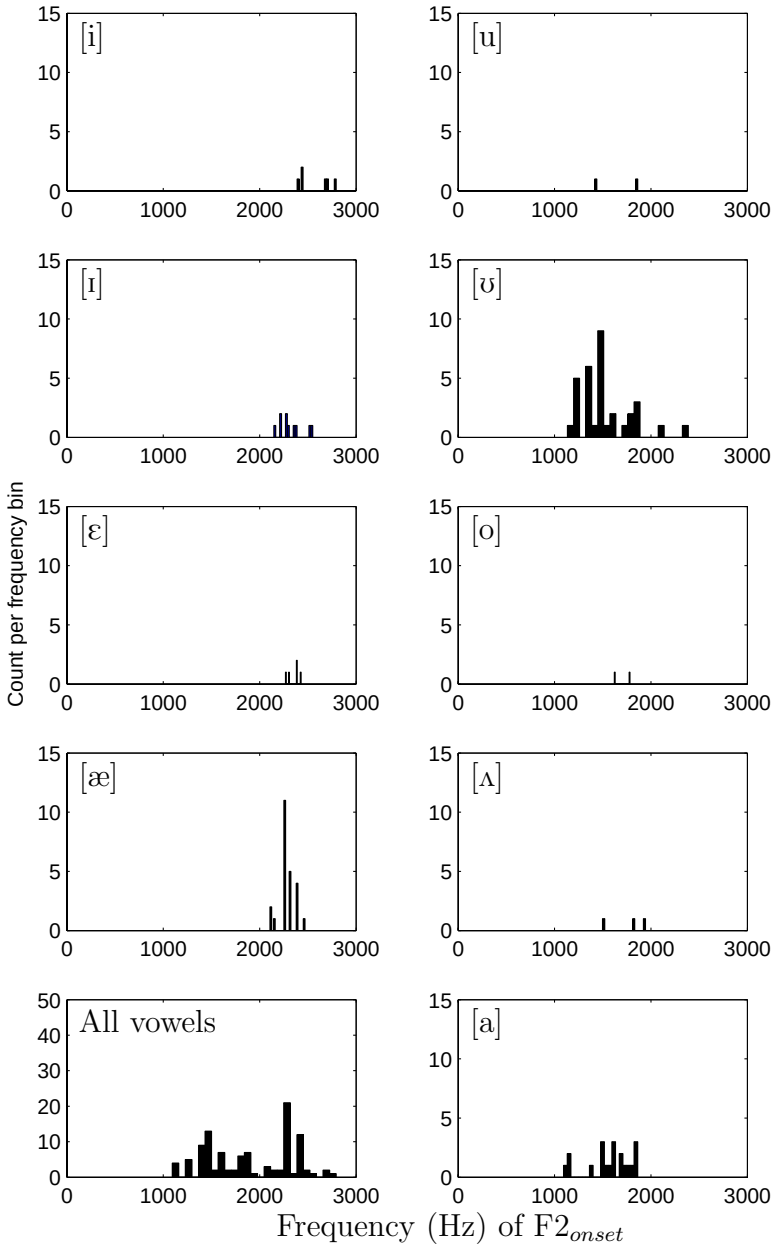


Figure 4-8: Same as in Figures 4-6 and 4-7 except for velars.

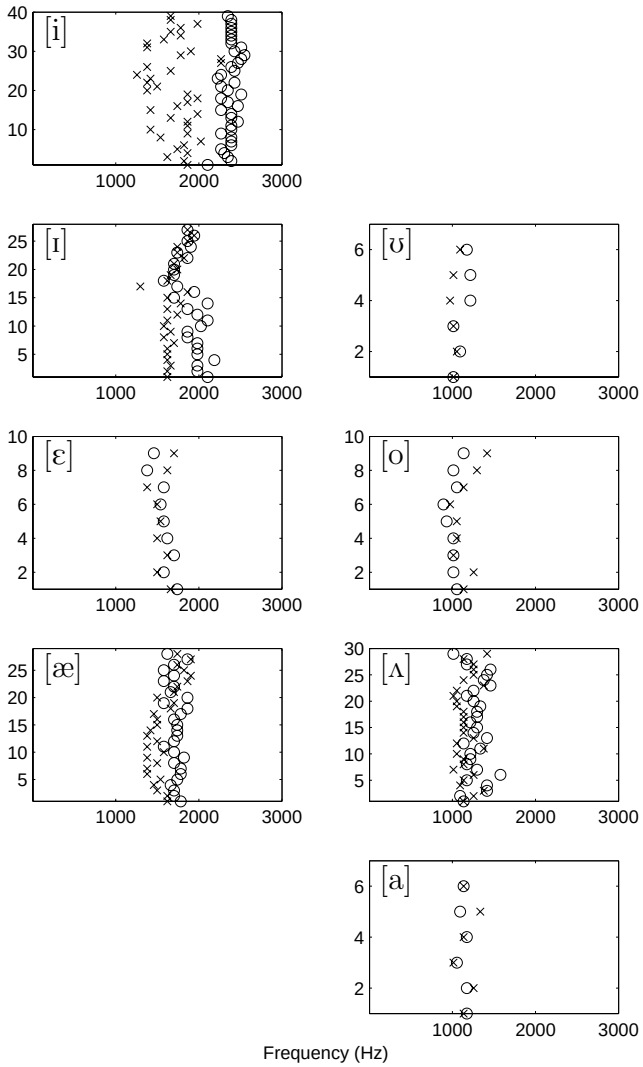


Figure 4-9: Scatter plot of $F2_{onset}$ ('x') and $F2_{vowel}$ ('o') for labials in all vowel contexts. 'x's and 'o's at the same level (y-axis) correspond to the $F2_{onset}$ and $F2_{vowel}$ measurements from the same CV transition. Most 'x'-'o' pairs either are located close together between lower airway resonances, or far enough apart to be on opposite sides of a lower airway resonance. In the back vowel contexts, all 'x'-'o' pairs are below the second lower airway resonance. In the front vowel contexts, 'x'-'o' pairs lie between the second and third lower airway resonances.

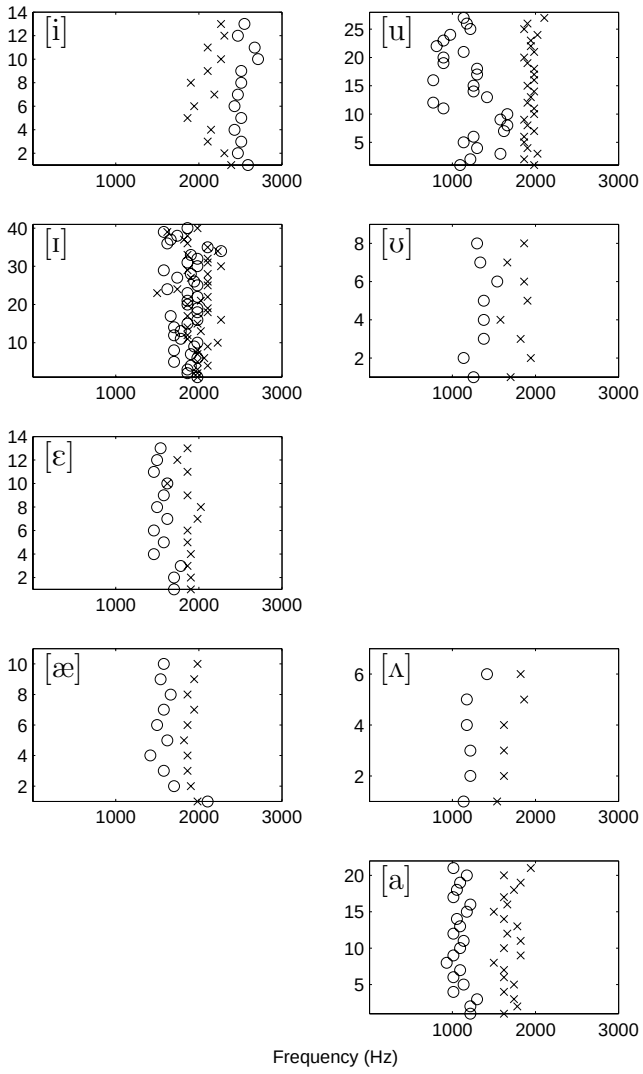


Figure 4-10: Same as in Figure 4-9 except for alveolars. In back vowel contexts the 'x'-'o' pairs lie on opposite sides of the second lower airway resonance. In front vowel contexts the 'x'-'o' pairs lie either between the second and third lower airway resonances.

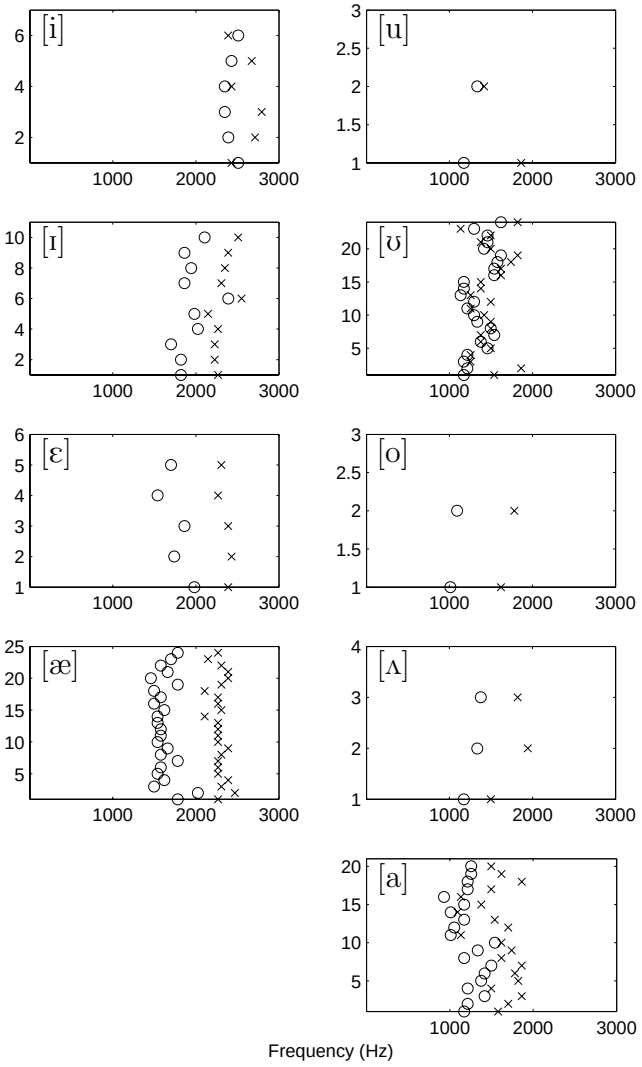


Figure 4-11: Same as in Figures 4-9 and 4-10 except for velars. In back vowel contexts the 'x'-'o' pairs tend to lie on opposite sides of the second lower airway resonance. in front vowel contexts the 'x'-'o' pairs either lie across the third lower airway resonance (e.g. [ɪ], [ɛ], [æ]) or above the third lower airway resonance (e.g. [i]).

and other formants.

- 3) If F2 remains between the second and third lower airway resonances, the consonant must be either an alveolar or a labial, and the vowel must be [-back]. The distinction between alveolar and labial place may be made by reference to F3 and other formants.

- 4) If F2 crosses from above the third lower airway resonance to below it, the consonant must be velar and the vowel must be [-back].

These generalizations may also be useful in speech perception. If a speaker's lower airway resonances can be detected, then generalizations like these may be used directly by a listener to decode the phonological/phonetic signal produced by the speaker. Speaker normalization might then be recast as a resonance detection problem rather than a mapping problem. Although this issue of speaker normalization will not be treated further, the question of whether the lower airway resonances play a role in speech perception is explored in chapter 5.

4.3.3 *Major conclusions of Chapter 4*

In an acoustic study of one speaker's speech, vowel F2 data fell into a bimodal distribution divided by the second lower airway resonance. Front vowels had a higher F2 frequency and back vowels had a lower F2 frequency. An estimate of the speaker's second lower airway resonance was within 20 Hz of the value calculated from the model in Chapter 3, and therefore the modeled frequencies of the second and third lower airway resonances were used for further comparisons with the speaker's second formant.

Transitions from a stop consonant to a vowel indicated that F2 onsets and endpoints in the transitions fell away from the lower airway resonances, while the transi-

tion often crossed one of the lower airway resonances. Furthermore, consonant places of articulation were partially determined by the relationship between F2 and the lower airway resonances.

In Chapter 5, the effects of the second lower airway resonance on the perception of a vowel-consonant F2 transition are explored.

Chapter 5

Lower Airway Resonances in the Perception of Vowels

This chapter contains data and selections from a paper submitted to the Journal of the Acoustical Society of America, in collaboration with Asaf Bachrach and Nicolas Malyska.

5.1 Method

5.1.1 *The stimuli: An overview*

A vowel [æ] was copy-synthesized from a natural recording made by an adult male, and simplified slightly in order to manipulate a single parameter (the frequency of the second lower airway resonance) systematically. Two versions of the synthetic vowel were made, in which the second subglottal resonance was either 1300 Hz or 1500 Hz. Two naturally produced utterances ('up there' and 'apter') were recorded ('apter' is the comparative form of 'apt'). The initial vowel in each was deleted and the synthetic vowel inserted in its place. There were thus 4 basic utterances. The

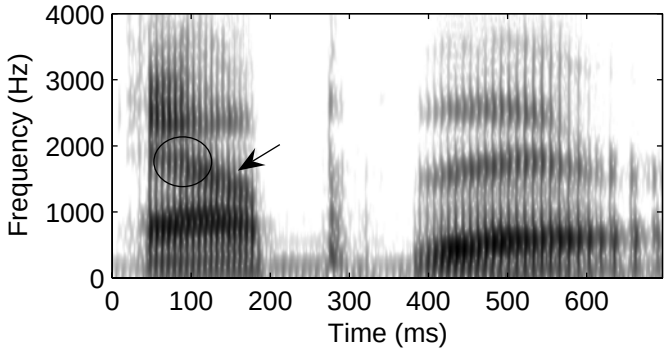


Figure 5-1: Spectrogram of the naturally produced vowel [æ] in the word 'apter'. The arrow indicates the interaction of the formant with the lower airway zero, and the oval marks the region of interaction between the formant and the lower airway pole.

utterances were then gated eight times (within the initial synthetic vowel) to form 32 stimuli, such that each stimulus contained some portion of the synthetic vowel followed by the naturally produced speech.

Natural utterances of the words 'upter' and 'up there' were used to create fillers. Each filler word was gated four times, yielding 8 fillers. The filler words began with the back vowel [ʌ] so that some words with long gates would be clearly perceived as back vowels, preventing subjects from relying on gate durations when making judgments.

5.1.2 *The synthetic vowel*

We copy-synthesized the initial vowel in 'apter' while making some reasonable simplifications. In the natural vowel, F2 began near 1800 Hz and fell to near 1200 Hz before the labial closure. In the synthetic vowel, F2 fell linearly from 1800 to 1200 Hz in 145 ms. In both the natural and synthetic vowels, F1 rose slightly over 110 ms and then fell toward the labial closure. F3 fell slightly over the duration of the natural vowel, and in the synthetic vowel it fell linearly from 2420 to 2400 Hz. Spectrograms of the copied and synthetic vowels are given in Figures 5-1 and 5-3, respectively.

In the natural vowel, the effects of the second lower airway resonance on F2 were apparent (cf. Figure 5-1). In the synthetic vowel, we simplified the pole-zero pair and replaced it with a zero. The elimination of the lower airway pole allowed us to control the stimuli and interpret the results with greater precision.

Since the synthesizer we used required pole-zero pairs ([37]), our zero was synthesized by a zero and a pole at the same frequency. The pole had a bandwidth of 500 Hz, and the zero had a bandwidth of 100 Hz. Our synthetic F2 had a bandwidth of 80 Hz, and F1 had a bandwidth of 40 Hz. Figure 5-2 shows the spectrum of the effective zero in relation to its component pole-zero parts, and in relation to F2 at the same frequency.

We synthesized the zero at two different frequencies, creating the two conditions for our experiment. In one condition the zero was synthesized at 1300 Hz, and in the other condition it was synthesized at 1500 Hz. Because F2 fell linearly from 1800 to 1200 Hz, it crossed the zero relatively late in the vowel in the 1300 Hz condition, and relatively early in the 1500 Hz condition. We hypothesized that the stimuli with a 1300 Hz zero would be perceived as containing an initial front vowel more often than the corresponding stimuli with a 1500 Hz zero.

5.1.3 *Gating*

The stimuli were gated 8 times in the synthetic vowel. Consecutive gates were separated by a single pitch period, and located at the waveform zero-crossing just before the initial rise at the beginning of the pitch period. The first gate was located at the beginning of the 7th pitch period from the end of the vowel; the eighth gate was located at the beginning of the 14th pitch period from the end of the vowel. There were 20 pitch periods in the whole vowel. The range of pitch periods that were gated were chosen so that they encompassed the region of the transition from perception of

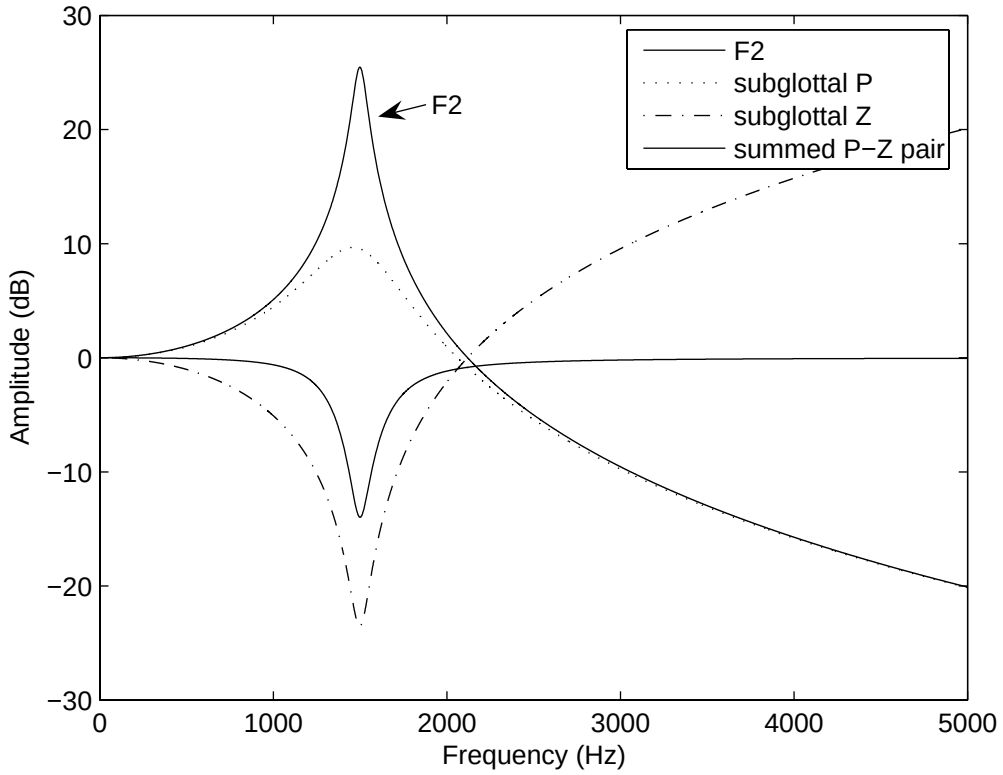


Figure 5-2: The effective zero was synthesized by placing a pole and a zero at the same frequency but with different bandwidths. Bandwidths were: 80 Hz for F2; 100 Hz for the zero; 500 Hz for the pole.

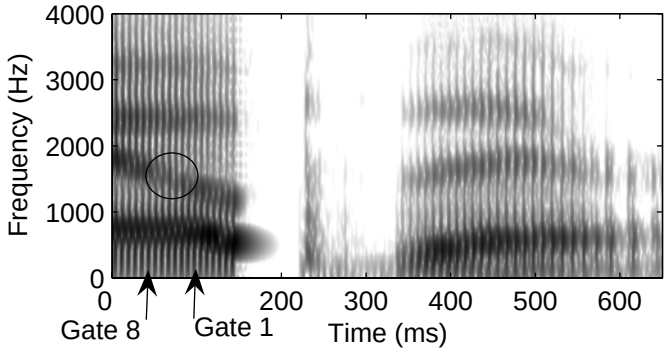


Figure 5-3: The structure of the stimuli (for the non-word 'ap there'). The initial vowel was synthetic, but the rest of the stimulus was naturally produced speech. The circle indicates the location of the visible effect of the zero on F2. The arrows indicate the locations of the first and eighth gates.

a front vowel to perception of a back vowel as less and less of the vowel was presented.

The filler words were gated 4 times in the same way as the stimuli, except there were two pitch periods between each consecutive gate. The first gate was located at the beginning of the 7th pitch period from the end of the vowel, and the fourth gate was located at the beginning of the 13th pitch period from the end of the vowel.

The fundamental frequency during the synthetic vowel was not constant, and hence individual pitch periods did not have the same duration. In the range of pitch periods that were gated, periods averaged 8 ms, and varied by less than 1 ms.

Figure 5-3 illustrates the structure of our gated stimuli. The longest gates (those with the earliest onsets) contain parts of the vowel in which the second formant crosses the effective zero; the shortest gates (those with the latest onsets) contain only those parts of the vowel in which the second formant does not cross the effective zero. Stimuli with the longest gates were expected to be perceived as containing an initial front vowel, and stimuli with the shortest gates were expected to be perceived as containing an initial back vowel. The number of the gates (1-8) is in order from the shortest gate to the longest gate.

5.1.4 *Procedure*

In an initial training phase of the experiment, the 32 stimuli and 8 fillers were presented to the subjects in random order with no feedback. Following a 30 second pause, the 32 stimuli and 8 fillers were presented 15 times each, in random order. Thus, there were 40 training stimuli and 560 experimental stimuli, yielding a total of 600 stimuli. After half of the 560 experimental stimuli were presented, there was another 30 second pause before continuing with the second half.

In each trial, the stimulus was presented twice with a 750 ms interval between the onset of each presentation. Subjects performed a two alternative forced choice lexical identification task in each trial. Of the two alternative choices, one was the real word or phrase 'apter' or 'up there', and the other was 'not "apter"' or 'not "up there"'. Responses were not timed, and subjects were informed at the beginning that they could pace themselves. Most subjects completed the experiment in about 30 minutes.

Subjects listened to the stimuli in a sound treated booth, wearing a pair of Cyber Acoustics HE-200 headphones. The volume was fixed by the experimenter at a moderately low level.

5.1.5 *Subjects*

Ten subjects participated in the experiment. All ten were native speakers of American English, and graduate students at MIT. They were naive to the purpose of the experiment. There were five males and five females.

5.2 Results

Figure 5-4 shows the results of the experiment averaged across all ten subjects, and across the two sets of stimuli ('apter' and 'up there'). The results for the 'apter'

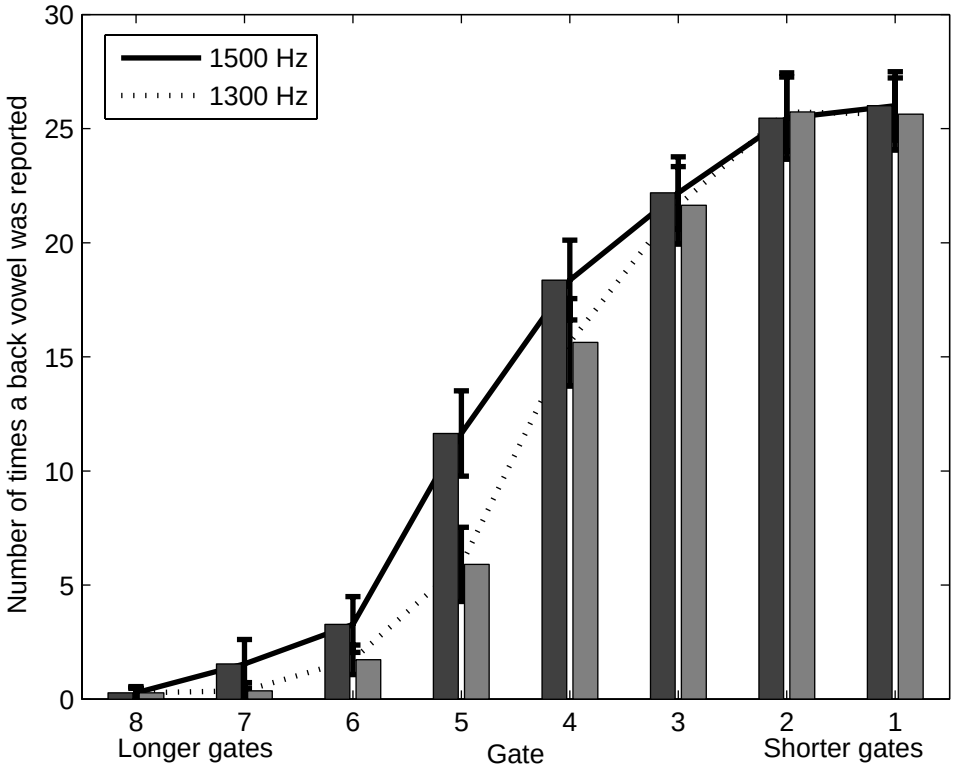


Figure 5-4: Identification curves with standard error for the 1300 and 1500 Hz conditions, averaged across subjects.

stimuli were similar to those for the 'up there' stimuli, indicating that any lexical bias effect on the vowel perception was insignificant. Results for male and female subjects were also similar. Plotted along the abscissa are the different gates, with longer gates toward the left and shorter gates toward the right. The percentage of stimuli reported to contain a back vowel are plotted along the ordinate. Thus, for the longest gates, few stimuli were reported to contain a back vowel, whereas most stimuli were reported to contain a back vowel for the shortest gates.

For the intermediate gates (gates 4-6), there is a significant difference (paired t-test, $p < 0.01$ for gate 5, $p < 0.02$ for gate 4, $p < 0.1$ for gate 6, $p > 0.1$ for all others)

between the 1300 and 1500 Hz conditions. In the 1500 Hz condition more stimuli were reported to contain a back vowel than in the 1300 Hz condition. Standard error bars are shown in order to illustrate the variation across subjects. Despite this variation, all subjects showed similar response patterns (Figure 5-5). For intermediate gates, each subject reported more words to contain a back vowel in the 1500 Hz condition than in the 1300 Hz condition. A point-biserial correlation coefficient of $r = 0.5$ was obtained for the 5th gate, in which the difference between the two conditions was greatest. For the 3rd and 7th gates, the point-biserial correlation coefficient was less than 0.25.

5.3 Discussion

The results indicate that the perception of vowel backness in the context of a spoken utterance can be altered as a result of manipulating the frequency of an effective acoustic zero in the vowel spectrum. Thus far we have followed [48] and [5] in suggesting that the boundary between front and back vowels is defined by the location of the second subglottal resonance. The results are consistent with this view. However, the presence of the effective acoustic zero in the vowel causes several changes to the overall spectrum of the vowel, including the apparent frequency and amplitude of the second formant, and the ratio of these parameters to those of the third formant. Thus, it is possible that the cause of the altered percept of vowel backness may be due to one of these other, secondary factors, rather than to the frequency of the acoustic zero per se. In this section, we will briefly review three alternative explanations for our results. Ultimately, none of these alternative explanations appears to be plausible.

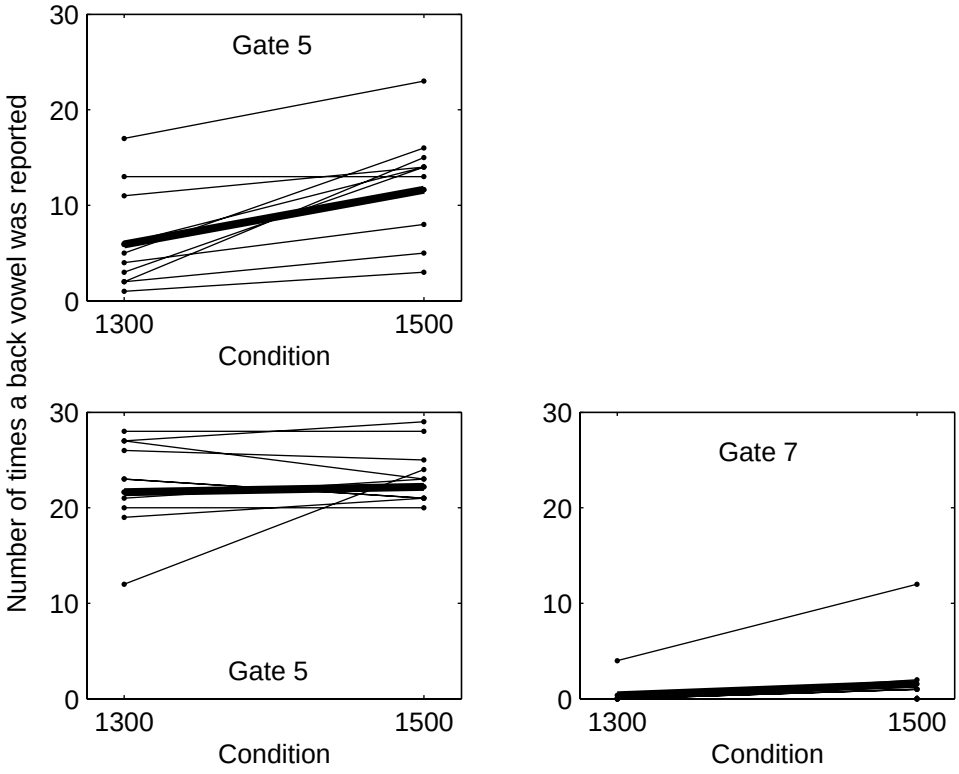


Figure 5-5: Individual subjects' responses to three pairs of stimuli (thick lines are averages). In gates 3 and 7 subject responses in the 1300 Hz and 1500 Hz conditions were similar; in gate 5 the responses differed, with more back vowels being reported in the 1500 Hz condition.

5.3.1 *Alternative #1: F3-F2 Bark difference*

Several studies have found that front and back vowels may be distinguished from each other by referring to the frequency difference between the second and third formants ([7, 53]). Specifically, if F2 and F3 are separated by less than roughly 3.5 Bark, the vowel may be classified as a front vowel; if F2 and F3 are separated by more than 3.5 Bark, the vowel may be classified as a back vowel. Since our stimuli had a descending F2 during the vowel, the difference between F2 and F3 began less than 3.5 Bark and ended greater than 3.5 Bark. It is interesting to note that the region in which the results of the 1300 Hz and 1500 Hz conditions differed was when F2 and F3 were roughly 3.5 Bark separated. This was also the region forming the boundary between front and back vowel percepts more generally. Thus, it may be the case that the 3.5 Bark separation between F2 and F3 forms at least a crude boundary between front and back vowels, which may be further defined with respect to the second subglottal resonance. Alternatively, it is also possible that the 3.5 Bark separation between F2 and F3 functions as a boundary between front and back vowels precisely because 3.5 Bark lower than F3 is roughly the location of the second subglottal resonance in natural speech.

Regardless of the fact that the transition region in the identification curves in Figure 5-4 coincides with a 3.5 Bark separation between F2 and F3, it is important to note that the percept of vowel backness was manipulated within this region. It is possible that this effect could be accounted for in terms of the separation between F2 and F3, since the actual frequency of the output peak amplitude near F2 is not identical to the input F2 value (cf. Figure 5-6). However, the differences between the 1300 and 1500 Hz conditions are very small (on the order of 1% or smaller), whereas the 3.5 Bark boundary hypothesis has never been demonstrated to this degree of precision. Therefore the 3.5 Bark boundary hypothesis would not predict the results

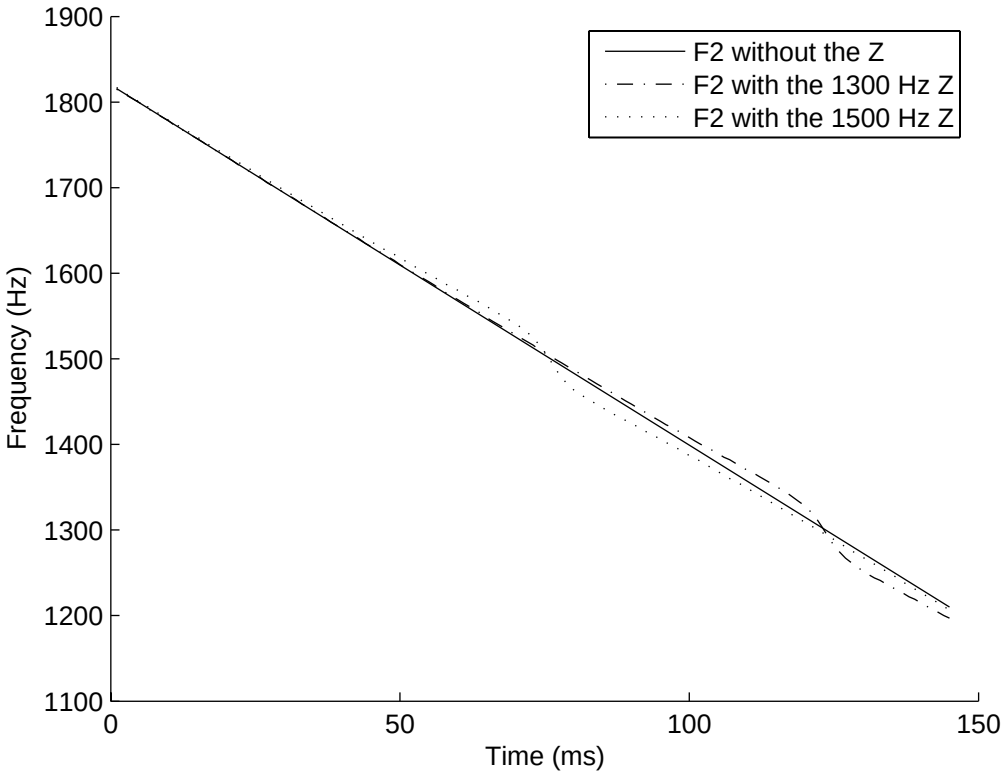


Figure 5-6: Input F2 and apparent F2 in the 1300 and 1500 Hz conditions. The thin solid line is the input F2, descending linearly. The dash-dot line is the apparent F2 in the 1300 Hz condition. The dotted line is the apparent F2 in the 1500 Hz condition.

reported here.

5.3.2 *Alternative #2: Center of gravity (F2')*

Closely related to the 3.5 Bark boundary hypothesis is the finding that closely spaced formants may be analyzed as a single excitation distribution by the auditory system, with a frequency-domain center of gravity dependant not only on the frequencies of the formants but also on their relative amplitudes ([7]). Center of gravity effects occur only when the formants in question are within about 3.5 Bark of each other,

close enough together that they may be integrated into a single distribution. Thus, if the lower formant has a relatively high amplitude, the center of gravity will be biased toward low frequencies; if the lower formant has a relatively low amplitude, the center of gravity will be biased toward high frequencies. If F2 and F3 are the two formants being integrated into a single center of gravity measure (F2'), vowels with a high F2 amplitude will be more likely to be perceived as back vowels, since the center of gravity is lower; vowels with a low F2 amplitude will be more likely to be perceived as front vowels, since the center of gravity is higher.

The effective acoustic zero introduced in the synthesis of our vowel stimuli affected the amplitudes of F2 and F3 (cf. Figure 5-7). The differences in F2 amplitude between the 1300 and 1500 Hz conditions approach 20 dB, whereas the difference in F3 amplitude is negligible. In the early part of the vowel ($<$ 100 ms), the F2 amplitudes in the two conditions would actually predict the opposite of the results we obtained, if the center of gravity effect were the cause. Here, the F2 amplitude is smaller in the 1500 Hz condition than in the 1300 Hz condition, which should lead to a higher center of gravity (F2') in the 1500 Hz condition, resulting in more front vowel reports. We found that there were more back vowel reports in the 1500 Hz condition than in the 1300 Hz condition. In the later part of the vowel ($>$ 100 ms), the formants are far enough apart that they should not be analyzed as a single distribution. The center of gravity hypothesis is therefore ruled out.

5.3.3 *Alternative #3: Spectral tilt*

Recent studies have explored the possibility that spectral tilt plays a role in distinguishing front from back vowels. Specifically, front vowels should have a smaller spectral tilt than back vowels, because a high second formant will boost the amplitudes of all higher formants. However, this effect has been demonstrated only in

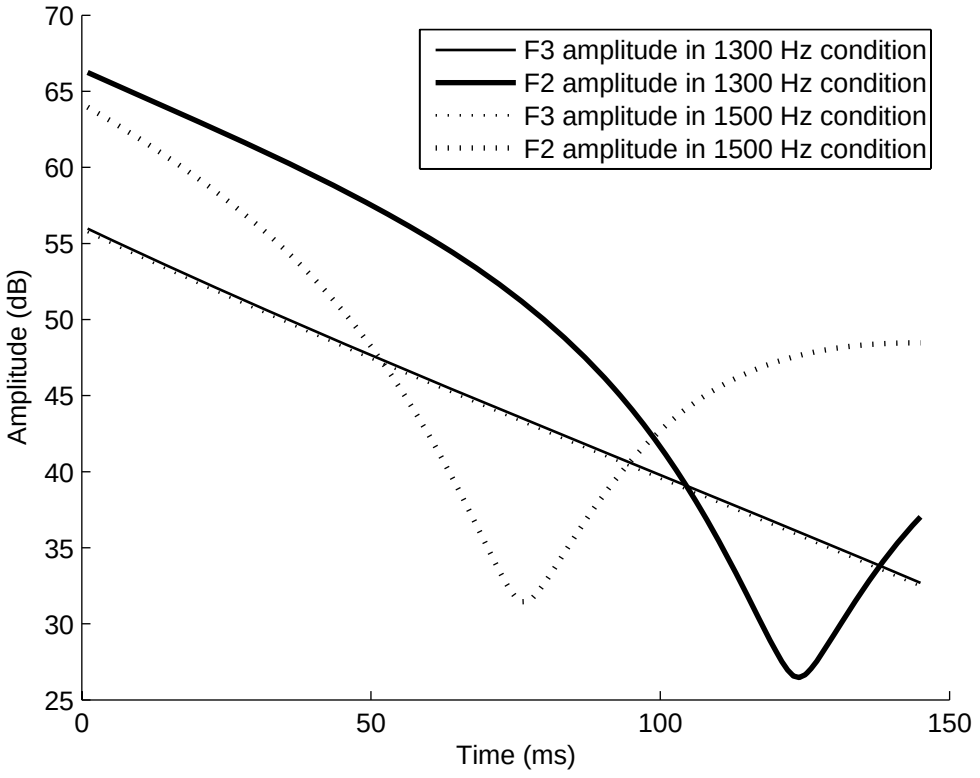


Figure 5-7: Amplitude of apparent F3 and F2 in the 1300 and 1500 Hz conditions. The solid lines are the amplitudes in the 1300 Hz condition; the dotted lines are the amplitudes in the 1500 Hz condition; The bold lines are the amplitudes of F2 in the two conditions; the light lines are the amplitudes of F3 in the two conditions.

steady-state vowels ([27]), and was not found for time-varying vowels ([36]). Since our synthetic vowel stimuli were time-varying, it is therefore unlikely that the spectral tilt hypothesis can account for the results.[1]

5.3.4 *Depth of the zero*

The effective zero used in the synthesis of our stimuli had a depth of roughly 14 dB. This depth is rather large. It should be noted, however, that zeros with depths as little as 2 dB have been reported to be perceptually salient [42] (cf. also [41, 15]). Furthermore, as illustrated in Figure **??**, the change in amplitude between a pole and a zero can be as much as 10 dB, even when their bandwidths are reasonably large. Therefore, although more work remains to be done, the depth of our effective zero is not likely a confound.

5.4 Some further remarks

Additional less formal experiments have been carried out to further investigate the role of the second lower airway resonance in the perception of vowel backness, as well as of stop place of articulation. These will be described briefly and qualitatively in this section. Further work in this area should systematize these or similar experiments and report quantitative data.

In the experiment described above, the zero was either 1300 Hz or 1500 Hz. Experiments have also been carried out in which the frequency of the zero was varied in steps of 50 Hz between 1200 Hz and 1700 Hz. When zero rises from 1200 to about 1500 or 1550 Hz, the boundary between front and back vowel percepts shifts as in Figure 5-4. As zero continues to rise to 1700 Hz, the boundary begins to shift

[1] I thank Terrance Nearey (p.c.) for pointing out this alternative, as well as suggesting the response adopted here.

back toward the condition with a low frequency zero. Presumably this is due to the physiological implausibility of such a high frequency zero.

The frequency above which zeros are considered to be physiologically implausible should, however, be dependent upon the frequency of F3 and other factors that cue the overall size of the speaker's vocal tract (and lower airway). In a separate experiment, it was found that by raising the frequency of F3, the frequency of the zero for which the perceptual boundary begins to shift back toward the condition with low zero is shifted upward. This experiment, however, was performed with a very small number of subjects and the resulting data were noisier than in the main experiment described above. This could be due to a difference in methodology in which several conditions of zero and F3 frequency were tested simultaneously, rather than only two zero conditions.

Finally, some very informal experimentation with copy-synthetic utterances indicated that the interaction of F2 with the second lower airway resonance helped to cue place of articulation in a following stop. A natural utterance of [pɛp] was copy-synthesized, but the result sounded like [pɛt]. In this utterance, F2 was rather steady throughout the vowel, though it did fall a little at the end. A zero at 1450 Hz was introduced so that the F2 crossed it just before the closure of the following stop. The result was not a completely disambiguated [pɛp], but several listeners could hear a difference, with the 1450 Hz zero condition sounding more like [pɛp] and the null condition sounding more like [pɛt].

5.4.1 *Major conclusions of Chapter 5*

Manipulation of the second lower airway resonance (represented as an effective zero) caused a shift in the perception of a front vs. back vowel in the context of a real word or phrase. This result could not be attributed to other factors, such as the effect of

the lower airway resonance on the Bark difference between F2 and F3, or the center of gravity of F2 and F3, or the tilt of the vowel spectrum. The perceptual effect of the second lower airway resonance, together with the acoustic effects and their relation to vowel categories (discussed in Chapters 3 and 4, and in Stevens [48] and Chi and Sonderegger [5]) is consistent with the view that the distinctive features are defined by non-linear, quantal relations between articulatory configurations (of the vocal tract) and acoustics (of the whole speech system, including the lower airway). Some implications for the quantal theory are discussed in Chapter 6.

Chapter 6

General Discussion

In the previous four chapters, a model of the lower airway was described and its basic properties characterized; the main effects of the lower airway acoustic properties on vowel spectra were characterized, including the observation that vowels contrasting in one (±[back]) feature might have a particular formant (F2) on opposite sides of a lower airway resonance; an acoustic study was carried out which indicated that discontinuities in F2 due to lower airway resonances are common and they may be used to characterize consonant place of articulation and vowel category in consonant-vowel transitions; and a perceptual study was carried out in which manipulations of a zero representing the second lower airway affected listeners' percepts of a vowel-consonant transition, specifically whether the vowel was front or back. These results are consistent with a view of distinctive feature theory and quantal theory which will be discussed in this chapter.

6.1 Quantal theory, landmark theory, and enhancement theory

Quantal theory [46, 47, 48] has largely been discussed in terms of spectral snapshots of a speech signal in time. The theory has been developing in recent years in conjunction with landmark theory [49] and enhancement theory [34, 35], which involve the time domain of a speech signal more explicitly. Briefly, this constellation of theories work together as follows: Landmarks are regions of the speech signal which cause the auditory system to respond robustly. Landmarks can be abrupt change in the speech signal, such as the sudden onset of frication or a stop release burst. The auditory system is sensitive to abrupt changes, and therefore the region around these landmarks contain some of the most salient information in the speech signal (information carried elsewhere in the speech signal will simply not cause the auditory system to respond so strongly). Alternatively, landmarks can be vowel nuclei, where the speech signal is loudest and therefore solicits a robust auditory response.

Regions near these landmarks are where the quantal cues to distinctive features are thought to be most salient. Enhancing acoustic cues (due to enhancing gestures) are thought to occur further away from the landmarks and occupy a longer stretch of the speech signal than the quantal cues do. In running speech, the quantal cues can disappear because of gestural overlap between adjacent segments, whereas the enhancing gestures are, perhaps, superimposed on the main articulatory gestures in such a way that they never disappear.

In general the quantal theory has been presented as a theory in the spectral domain. In fact, the precursor to the current Lexical Access from Features (LAFF) project was called Lexical Access from Spectra (LAFS), and both of these projects have been dependent upon quantal theory as a primary guide for developing speech

synthesis and recognition technologies. The familiar graphs that are used in presentations of quantal theory, or the quantal nature of individual features, are usually spectral - that is, for a specific articulatory configuration, a certain spectral F-pattern (to use Fant's terminology [13]) is obtained. Time does not play an explicit role in any presentation of quantal theory, and in fact it is generally recognized that quantal theory has little or nothing to say (currently) about temporal contrasts in phonology, such as between long and short vowels or singleton and geminate consonants.

In the next section, a different perspective on quantal theory will be outlined, incorporating the data presented in the previous chapters.

6.2 A new perspective on quantal theory

Just as every signal can be represented in either the time or the frequency domain (the two domains are related via, e.g. a Fourier transform), so too the quantal articulatory-acoustic characterization of speech sounds and distinctive features can be represented spectrally or temporally. In this section, the spectral and the temporal domains will be combined (similar to the familiar spectrographic domain).

For instance, consider two segments adjacent in time, S_1S_2, which contrast only in one feature [±F]. Let S_1 be [+F] and S_2 be [-F]. If [±F] is quantally defined, it will normally be represented in an articulatory-acoustic space in which one articulatory parameter varies along the x-axis and the acoustic properties (usually the F-pattern) of the resulting articulatory configurations are represented along the y-axis. As the defining articulator (e.g. the tongue tip) moves from a position in the [+F] part of the quantal curve toward the [-F] part of the quantal curve, assuming that the movement is continuous and smooth, a sudden change in the acoustics will occur as the quantal boundary is crossed. This sudden change is analogous to the landmarks discussed above.

In fact, the kinds of landmarks discussed above which result from consonant onsets and offsets can always be thought of as a result of the spectrographic change from one value of a feature to its opposite. The sudden onset of frication in a sequence such as [æʃ], for instance, is a sudden change in the excitation of the vocal tract (from periodic excitation to noisy excitation), that is, a change in the feature [±son]. The concomitant sudden change in the spectral energy at low frequencies is a result of the change in the location of the excitation source, to the alveo-palatal region of the mouth, such that zeros are introduced that reduce the energy in the lower frequency band. This is equivalent to a change in the feature [±cons]. Similarly, a change in the place of articulation occurs, and is reflected in a change in the F-pattern (F2 rises quickly into the [ʃ], and F3 also rises somewhat). The sudden change in the acoustics as the place of articulation is altered - that is, the landmark - is in this case the discontinuity in F2 as it passes through the third lower airway resonance.

This view of landmarks is somewhat different from the view presented in Stevens [49]. In that exposition, for instance, a discontinuity in a formant track would not be considered a landmark. Conversely, under the present exposition the consonantal landmarks of Stevens [49] are merely one class of landmarks. Different classes of landmarks can be distinguished in several ways. One could speak of landmarks resulting from changes in source excitation type or location; or landmarks resulting from changes in place of articulation or the sudden opening of side branches (such as the nasal or sublingual cavities). Alternatively, landmarks can be classified according to the width of the frequency band over which they occur. Stop bursts, for instance, are wide band landmarks, whereas discontinuities in F2 due to lower airway resonances are narrow band landmarks.

In the presentation of such narrow band landmarks in chapters 4 and 5, an emphasis was placed on the role of these landmarks in the marking of place of articulation changes. Both the occurrence of the landmark and its directionality (e.g. from F2

higher than the second lower airway resonance to F2 below it, or vice versa) were important. This view of the role of landmarks - which is rooted in quantal theory - leads to a somewhat different way of viewing the speech signal and how it relates to phonological units (features, segments, etc). For instance, it is compatible with a window model of speech along the lines of Keating [33]. F2 need not reach a particular frequency in order to signal a particular place of articulation; it need only cross an appropriate lower airway resonance such that a discontinuity in the F2 track appears.

Furthermore, a landmark resulting from a change in one feature need not occur simultaneously with a landmark resulting from a change in another feature. For instance, in the sequence [da], F2 begins above the second lower airway resonance and falls to below it. The landmark resulting from this interaction of F2 and the lower airway resonance usually occurs some time after the composite landmark caused by the change in the features [cont], [cons], and [son]. On the other hand, some landmarks, such as the composite landmark just mentioned, always occur simultaneously. The composite landmark in [da] consists of three separate landmarks: the change in excitation source, the change in source location, and the change in radiation type ([cont] and [nasal] might be thought of as radiation features, since they determine whether sound is radiated through the mouth, the nose, both, or neither). If the temporal alignment of different features belonging to a single segment is not restricted to be simultaneous, this raises several possibilities for the integration of quantal theory with articulatory phonology. For instance, the relative synchrony or asynchrony of landmarks for different features might relate to the relative phase differences between gestures in an articulatory score. Furthermore, the asynchrony of landmarks has already been discussed in the context of automatic speech recognition technology [40], and a further systematic study of these asynchronies could be of use in developing more robust front-ends for ASR systems, removing some of the burden which higher-level language models currently carry.

6.3 Different ways to classify distinctive features

In feature geometry distinctive features are organized in a hierarchy that encodes the relations between different features. For example, the features [anterior] and [distributed] are cosubordinate to the coronal place node, according to Halle [20], indicating that they both are affected by assimilation rules involving the place nodes. Feature geometry can take several forms, however, depending on how one decides to classify the relations between features. For instance, one can define feature relations on the basis of purely phonological patterns, or on the basis of purely phonetic characteristics, or both. Halle [20] makes use of the phonetic distinction 'articulator-bound' vs. 'articulator-free' to justify a geometry that also has phonological significance. The International Phonetic Association relies on a similar distinction between 'place' and 'manner' features. Halle's articulator-bound features are equivalent to the IPA's place features, and Halle's articulator-free features are equivalent to the IPA's manner features.

Another way to classify features phonetically is by referring to them as source features or filter features, as Jakobson, Fant, and Halle [30] did. A fourth possibility is to distinguish features according to how wide-band the landmarks associated with them are. In general, manner, source, or articulator-free features will be associated with wider-band landmarks than place, filter, or articulator-bound features. It would be interesting to explore this classification scheme further, but it is beyond the scope of this thesis to do so here.

6.4 Direct and indirect cues to features

The articulator-bound, or place, features are associated with landmarks caused by discontinuities in the formants due to the interaction of the formants with lower

airway resonances. Since the lower airway resonances are always present and relatively constant, whether near a formant or not, listeners may be able to keep track of their frequencies for a given speaker, after they have been determined from a brief or lengthy sample of that speaker's speech. If a formant moves in the direction of a lower airway resonance but does not form a landmark (for instance, if the formant disappears because of a change in [cons] before it crosses the lower airway resonance), the listener could extrapolate the movement of the formant and determine that it would have crossed a lower airway resonance (shortly after its disappearance). Although the landmark is not present, there are cues in the signal (such as the movement of the formant and the fixed location of the lower airway resonance) that a listener can use to infer the features and feature changes involved. A landmark is merely a relatively direct cue vs. the movement of a formant toward a lower airway resonance (and the landmark is potentially more salient).

This characterization of more and less direct cues (and landmarks specifically as more direct cues) presents a slightly different perspective on the relationship between defining (i.e. quantal) and enhancing gestures and acoustics. Stop bursts as in [ta], for instance, are generally considered to be defining acoustics for place of articulation of a stop, whereas the specific properties of formant transitions are due to enhancement (e.g. whether they are short and quick, long and drawn out, or whether they are noisy as in the case of an aspirated stop release). Under the perspective presented here, stop bursts are landmarks associated not with place of articulation but with manner of articulation (or source excitation type, or radiation type); they are direct cues to the change in the features [cons], [cont], and/or [son]; the nearly simultaneous landmark as a formant crosses a lower airway resonance is similarly a direct cue to the change in place of articulation.

Features do not change at the rate of segments, however, and therefore between one segment and the next at least one feature is likely to remain unchanged. In

this case, no landmark associated with that feature will occur. Thus, there are three possibilities with regard to the presence of a landmark: 1) absence of the landmark because its associated feature has not changed value, 2) presence of the landmark with the transition from the '+' value to the '-' value of the feature, and 3) presence of the landmark with the transition from the '-' value to the '+' value of the feature. (The second and third possibilities arise from the fact that landmarks are directional. For instance, a change from a [-cont] to a [+cont] consonant involves an acoustic change from silence to noise, whereas the reverse transition involves the reverse acoustic change.) Moreover, there are two possible conditions in which a landmark may be absent: a) when the feature is '+' valued, and b) when the feature is '-' valued.

It might be hypothesized that segment transitions which involve landmarks due to feature value changes are diachronically more stable since the presence of the landmark is presumably more salient than its absence. Investigation of this hypothesis is, however, beyond the scope of this thesis.

6.5 Speaker normalization

The role of lower airway resonances in speech production and perception may also have implications for speaker normalization. Generally formant frequencies (and the fundamental frequency) are normalized with respect to each other in order to produce some acoustic space in which normalized formants for all speakers lie close together. It is necessary for some sort of normalization to occur, since the absolute formant frequencies of an adult male, for instance, are much different from those of a child or an adult female. However, if lower airway resonances can be automatically detected, they may be useful in speaker normalization, since they are relatively fixed for any individual and the formant frequencies relative to these resonances play an important role in defining place of articulation features for vowels and consonants. It is possible

that the formants need be normalized relative to each other, but only relative to the lower airway resonances. If true, the lower airway resonances would provide a simpler and more natural means of normalization than current normalization paradigms.

Chapter 7

Summary, Conclusions, and Future Directions

7.1 Chapter by chapter summary

In the preceding chapters, the following have been accomplished:

In the second chapter, a model of the lower airway was constructed and analyzed. It was used to analyze measurements that have been made on humans both directly (e.g. via tracheostoma) or indirectly (e.g. via spectrographic analysis). The effects of varying the symmetry, the wall properties, the peripheral load impedance, and the number of generations of the lower airway were explored. Finally, the second and third lower airway resonances were found to be affiliated primarily with the left bronchial tree, and the potential utility in medicine of characterizing cavity-resonance affiliations more fully was broached.

In the third chapter, a model of the vocal tract coupled with the lower airway by means of a variable glottal impedance was constructed, and the effects of lower airway coupling on vowel spectra were characterized. The low back vowels were found to be

the most sensitive to changes in glottal impedance, due to their constricted pharyngeal cavity and the resulting closer match between the pharyngeal and glottal impedances. It has been suggested by Stevens [48] that the first lower airway resonance divides [+low] from [-low] vowels, and that hypothesis appears to hold under these modeling conditions. It was also suggested that the second lower airway resonance divides [+back] from [-back] vowels, and that the third lower airway resonance divides [+ATR] from [-ATR] front vowels. Finally, it was also suggested that the interaction of F2 with the first lower airway resonance may also divide [+ATR] from [-ATR] back vowels (F2 for [+ATR] back vowels is near 1000 Hz, which is the frequency that appears to be boosted in [a] and [ɔ] when the glottal impedance is small). The potential applications of these findings toward better understanding whispered speech and other non-modal voice qualities and pathologies was also suggested.

In the fourth chapter, speech from one speaker from a database collected by Black and colleagues [38] was analyzed. CV transitions showed that the onset F2 immediately after the release of a stop consonant appears to occur in specific frequency bands between lower airway resonances, depending on the stop place of articulation. Similarly, F2 measured in the steady-state portion of the vowel showed a pattern similar to that predicted in chapter 3 from the model. Five patterns of F2 transitions in a CV utterance were identified and characterized in terms of consonant place of articulation and vowel backness, and the potential utility of these characterizations in speech technologies and theories of speech perception was pointed out.

In the fifth chapter, a speech perception experiment was carried out (in collaboration with Asaf Bachrach and Nicolas Malyska) in which the effect on vowel perception of varying the frequency of the second lower airway resonance was studied. Given an otherwise identical vowel stimulus, listeners identified the vowel as [+back] more often when the second lower airway resonance was at a relatively high frequency than when it was at a relatively low frequency. The F3-F2 Bark hypothesis, the center of

gravity (F2') hypothesis, and the spectral tilt hypothesis could not account for the results, indicating that the effect was due directly to the frequency of the resonance.

In the sixth chapter, some implications of the results of the previous four chapters were discussed and related to issues in quantal theory, landmark theory, distinctive features, and speaker normalization. Specifically, a new perspective on quantal theory which incorporates the temporal domain was presented, with some implications for speaker normalization and for phonology.

7.2 Conclusions

This thesis presents theoretical evidence from models and experimental evidence from speech production and speech perception that lower airway resonances play a role in defining vowel feature contrasts, and perhaps more. The lower airway resonances appear to be actively used by listeners to aid in the perception of speech. Whether they are actively manipulated - or rather, whether the formants are actively manipulated relative to the lower airway resonances - by speakers is unknown. The possibility that the lower airway resonances play an important role in speech production and perception raises a large number of questions.

7.3 Future directions

The nature and extent of the role of lower airway resonances in speech production and perception, and in defining vowel and consonant feature contrasts, must be clarified more precisely. Future studies should explore more systematically the perceptual effects of representing the second lower airway resonance as a zero unpaired with a pole; of varying the bandwidth of the zero; of representing the resonance as a pole-zero pair and varying their bandwidths; of varying the frequencies of the pole and zero,

and the nonlinear interaction of the second formant with the pole. Similar studies with respect to the first and third lower airway resonances should be carried out. The sensitivity of F1 and F2 in [a] and [ɔ] to glottal impedance variations should be clarified, and the possibility of relating this sensitivity to measures of vocal quality and (dis)function should be explored.

Of further interest would be a fuller understanding of the resonance-cavity affiliations of the lower airway, such that lung health could be assessed noninvasively on the basis of a patient's speech. Finally, further studies in the biology and physiology of speech production should be carried out. Perhaps the distribution of pulmonary stretch receptors is more dense near the nodes or antinodes of the lower airway resonances, leading to a kind of sensorineural feedback mechanism for the control of speech. Or perhaps the interaction of vocal tract resonances (formants) with lower airway resonances is of broader interest in animal communication. Many kinds of animal calls consist of rising or falling chirps - perhaps the endpoints of some of these chirps are at frequencies near the animals' lower airway or tracheal resonances.

Appendix A

MATLAB code for the lower and upper airway models

```
function varargout = SGModel5_orig(varargin);
%
%
% by Steven M. Lulich
% lulich@speech.mit.edu

if ~nargin,
    action = 'new';
else,
    action = varargin{1};
end;

switch action,
```

```
case 'getTvt',      % Transfer function from glottis to lips
    f = varargin{2};
    w = 2*pi*f;
    dims = varargin{3};
    glottalDims = varargin{4};
    Zsg = varargin{5};
    Ug = 1;
    Us = 1;
    Zg = SGModel5_orig('getZg',w,glottalDims,Ug);
    Tvt = SGModel5_orig('Cramer',w,dims,Us,Zsg,Zg);
    varargout = {Tvt};
case 'Cramer',
    w = varargin{2};
    dims = varargin{3};
    Us = varargin{4};
    Zsg = varargin{5};
    Zg = varargin{6};

    constants = SGModel5_orig('getConstants');
    TsPerSection = [];
    TsCumSections = [];
    sectionDims = [];
    for j = 1:length(dims(1,:)),
        TsPerSection = [TsPerSection ceil(dims(1,j)/...
            constants.threshold)];
        TsCumSections = [TsCumSections sum(TsPerSection)];
        for k = 1:floor(dims(1,j)/constants.threshold),
```

```
        sectionDims = [[sectionDims] ...
            [constants.threshold;dims(2,j)]];
    end;
    if (dims(1,j)-constants.threshold*floor(dims(1,j)/...
        constants.threshold)) > 0,
        sectionDims = [[sectionDims] [dims(1,j)-...
            constants.threshold*floor(dims(1,j)/...
            constants.threshold);dims(2,j)]];
    end;
end;
%numberOfTSections = sum(TsPerSection);
numberOfTSections = length(sectionDims(1,:));

for f = 1:length(w),
    % Initialize the matrix
    for j = 1:2*numberOfTSections+3, % i.e., # of variables
        for k = 1:2*numberOfTSections+3,
            M(j,k) = 0;
        end;
        A(j) = 0;
    end;
    % Fill in the matrix for the normal node equations
    for j = 1:numberOfTSections, % i.e., # of nodes
        M(j,j) = 1;
        M(j,j+1) = -1;
        M(j,j+numberOfTSections+3) = 1;
    end;
```

```
% Fill in the matrix for the extra 2 node equations
j = numberOfTSections+1;
M(j,j) = 1;
M(j,j+1) = -1;
A(j) = Us;
j = numberOfTSections+2;
M(j,j) = 1;
M(j,j+1) = -1;
A(j) = -Us;

% Fill in the matrix from the normal loop equations
% First loop (at the lips)
j = 1;
[Sj SjMinus1 Hj HjMinus1] = SGModel5_orig('getZ2',j,...
    sectionDims,f);
M(numberOfTSections+j+3-1,j) = Sj + SjMinus1;
M(numberOfTSections+j+3-1,numberOfTSections+j+3) = -Hj;
% Other loops besides the one circling the subglottal
% and glottal impedances
for j = 2:numberOfTSections, % i.e., # of loops except the one
                                % at the lips
    [Sj SjMinus1 Hj HjMinus1] = SGModel5_orig('getZ2',j,...
        sectionDims,f); % two shunts plus two series impedances
    M(numberOfTSections+j+3-1,j) = Sj + SjMinus1;
    M(numberOfTSections+j+3-1,numberOfTSections+j+3-1) = HjMinus1;
    M(numberOfTSections+j+3-1,numberOfTSections+j+3) = -Hj;
end;
```

```
        % The extra loop circling the subglottal and glottal impedances
        j = numberOfTSections+1;
        [Sj SjMinus1 Hj HjMinus1] = SGModel5_orig('getZ2',j,sectionDims,f);
        M(numberOfTSections+j+3-1,j) = SjMinus1;
        M(numberOfTSections+j+3-1,numberOfTSections+j+3-1) = HjMinus1;
        M(numberOfTSections+j+3-1,j+1) = Zg(f);
        M(numberOfTSections+j+3-1,j+2) = Zsg(f);

        N = M;
        N(:,1) = A';
        T(f) = det(N)/det(M);
    end;
    varargout = {T};
case 'getZ2',
    w = 2*pi*varargin{4};
    sectionDims = varargin{3};
    j = varargin{2};

    constants = SGModel5_orig('getConstants');

    if j<=length(sectionDims(1,:)),
        if j>1,
            R1 = sqrt(sectionDims(2,j-1)/pi);
            elements1 = SGModel5_orig('getElements',w,constants,...
                sectionDims(1,j-1),R1,0.5,0);
            R2 = sqrt(sectionDims(2,j)/pi);
            elements2 = SGModel5_orig('getElements',w,constants,...
```

```
            sectionDims(1,j),R2,0.5,0);

        HjMinus1 = 1./(elements1.Ga + i*w*elements1.Ca);
        Hj = 1./(elements2.Ga + i*w*elements2.Ca);

        SjMinus1 = (i*w*elements1.La + elements1.Ra)/2;
        Sj = (i*w*elements2.La + elements2.Ra)/2;
    else,
        R2 = sqrt(sectionDims(2,j)/pi);
        elements2 = SGModel5_orig('getElements',w,constants,...
            sectionDims(1,j),R2,0.5,0);  % back cavity except last segment
        Hj = 1./(elements2.Ga + i*w*elements2.Ca);
        Sj = (i*w*elements2.La + elements2.Ra)/2;
        Zr = 0;
        SjMinus1 = Zr;
        HjMinus1 = 0;
    end;
else,
    R1 = sqrt(sectionDims(2,j-1)/pi);
    elements1 = SGModel5_orig('getElements',w,constants,...
        sectionDims(1,j-1),R1,0.5,0);
    HjMinus1 = 1./(elements1.Ga + i*w*elements1.Ca);
    SjMinus1 = (i*w*elements1.La + elements1.Ra)/2;
    Hj = 0;
    Sj = 0;
end;
varargout = {Sj,SjMinus1,Hj,HjMinus1};
```

```
case 'getZsg',
    w = 2*pi*varargin{2};
    networktype = varargin{3};
    generations = varargin{4};
    parameters = SGModel5_orig('getSGParams');
    % load('WodickaParams.mat');
    % parameters = p;
    [L, R, T, C] = SGModel5_orig('asymmtrach',parameters,generations);
    %main loop
    for j=generations:-1:1,
        for m=1:2^(j-1),
            if j==generations,
                ZL = SGModel5_orig('getDistalLoadImpedance');
            else,
                Z1 = Z{j+1,2*m-1};
                Z2 = Z{j+1,2*m};
                ZL = Z1.*Z2./(Z1+Z2); % equivalent to Z1/2 or Z2/2 if
                                      % the bronchial tree is symmetrical
            end;
            Z{j,m} = SGModel5_orig('recursiveTube',w,networktype,ZL,...
                1.1*L(j,m),R(j,m),T(j,m),C(j,m));
        end;
    end;
    varargout = {Z{1,1}};
case 'getZvt',
    w = 2*pi*varargin{2};
    dims = varargin{3};
```

```
    networktype = 1;
    Zr = 0;
    Z = Zr;
    for j = 1:length(dims(1,:)),
        L = dims(1,j);
        A = dims(2,j);
        Z = SGModel5_orig('recursiveTube',w,networktype,Z,L,...
            sqrt(A/pi),0.5,0);
    end;
    varargout = {Z};
case 'getUvt',
    f = varargin{2};
    w = 2*pi*f;
    dims = varargin{3};
    d = varargin{4};
    Zsg = varargin{5};
    Us = 1;
    Zvt = SGModel5_orig('getZvt',f,dims);
    Zg = SGModel5_orig('getZg',w,d,Us);
    Uvt = (Zg+eps)./(Zg + Zsg + Zvt)*Us;
    varargout = {Uvt};
case 'recursiveTube',
    w = varargin{2};
    networktype = varargin{3};
    ZL = varargin{4};
    L = varargin{5};
    R = varargin{6};
```

```
    T = varargin{7};
    C = varargin{8};
    constants = SGModel5_orig('getConstants');
    Z = eval(['SGModel5_orig(''network' num2str(networktype) ''',...
        w,ZL,constants,L,R,T,C);']);
    varargout = {Z};
case 'asymmtrach',
    parameters = varargin{2};
    generations = varargin{3};
    depths = parameters(:,1);
    lengths = parameters(:,2);
    radii = parameters(:,3);
    thickness = parameters(:,4);
    cfrac = parameters(:,5);
    dn = parameters(:,6);
    for j=1:generations
        for k=1:2:2^(j-1),
            if j==1,
                structure(j,:)=0;
                n = structure(j,k) + 1;
                newl(j,k) = lengths(n);
                newr(j,k) = radii(n);
                newh(j,k) = thickness(n);
                newcfrac(j,k) = cfrac(n);
            else
                structure(j,k)=structure(j-1,(k+1)/2)+1;
                n = structure(j,k) + 1;
```

```
                newl(j,k) = lengths(n);
                newr(j,k) = radii(n);
                newh(j,k) = thickness(n);
                newcfrac(j,k) = cfrac(n);

                structure(j,k+1)=structure(j,k)+dn(structure(j,k));
                n = structure(j,k+1) + 1;
                newl(j,k+1) = lengths(n);
                newr(j,k+1) = radii(n);
                newh(j,k+1) = thickness(n);
                newcfrac(j,k+1) = cfrac(n);
            end;
        end;
    end;
    varargout = {newl,newr,newh,newcfrac}; % = {L,R,T,C};
case 'network1',    % T-network with rigid walls
    w = varargin{2};
    ZL = varargin{3};
    constants = varargin{4};
    L = varargin{5};
    R = varargin{6};
    T = varargin{7};
    C = varargin{8};
    q = ceil(L/constants.threshold);
    elements = SGModel5_orig('getElements',w,constants,...
        constants.threshold,R,T,C);
    Zshunt = 1./(elements.Ga + i*w*elements.Ca);
```

```
    Zseries = (i*w*elements.La + elements.Ra)/2;
    for v = 1:q,
        if v == q,
            Lmod = L-(q-1)*constants.threshold;
            elements = SGModel5_orig('getElements',w,constants,Lmod,R,T,C);
            Zshunt = 1./(elements.Ga + i*w*elements.Ca);
            Zseries = (i*w*elements.La + elements.Ra)/2;
        end;
        ZL = Zseries + 1./(1./Zshunt + 1./(Zseries + ZL));
    end;
    varargout = {ZL};
case 'network2',    % T-network with yielding walls in parallel
    w = varargin{2};
    ZL = varargin{3};
    constants = varargin{4};
    L = varargin{5};
    R = varargin{6};
    T = varargin{7};
    C = varargin{8};
    q = ceil(L/constants.threshold);
    elements = SGModel5_orig('getElements',w,constants,...
        constants.threshold,R,T,C);
    Zshunt = 1./(elements.Ga + i*w*elements.Ca + ...
        1./(i*w*elements.Lwc + elements.Rwc + ...
        1./(i*w*elements.Cwc+eps)) + ...
        1./(i*w*elements.Lws + elements.Rws + ...
        1./(i*w*elements.Cws+eps)+eps)+eps); % why the extra eps's?
```

```
    Zseries = (i*w*elements.La + elements.Ra)/2;
    for v = 1:q,
        if v == q,
            Lmod = L-(q-1)*constants.threshold;
            elements = SGModel5_orig('getElements',w,...
                constants,Lmod,R,T,C);
            Zshunt = 1./(elements.Ga + i*w*elements.Ca + ...
                1./(i*w*elements.Lwc + elements.Rwc + ...
                1./(i*w*elements.Cwc+eps)) + ...
                1./(i*w*elements.Lws + elements.Rws + ...
                1./(i*w*elements.Cws+eps)+eps)+eps)+eps); % the extra eps's?
            Zseries = (i*w*elements.La + elements.Ra)/2;
        end;
        ZL = Zseries + 1./(1./Zshunt + 1./(Zseries + ZL));
    end;
    varargout = {ZL};
case 'getConstants',
    constants.rho = 1.14*10^-3;
    constants.eta = 1.86*10^-4;
    constants.nu = 1.4;
    constants.kappa = 0.064*10^-3;
    constants.cp = 0.24;
    constants.c = 3.54*10^4;
    constants.threshold = 1.0;
    constants.rhows = 1.06;
    constants.etaws = 1.6*10^3;
    constants.Ews = 0.392*10^6;
```

```
    constants.rhowc = 1.14;
    constants.etawc = 180.0*10^3;
    constants.Ewc = 44.0*10^6;
    varargout = {constants};
case 'getElements',
    w = varargin{2};
    constants = varargin{3};
    L = varargin{4};
    R = varargin{5};
    T = varargin{6};
    C = varargin{7};
    A = pi*R^2;
    elements.Ra = 2*L/(pi*R^3)*sqrt(w*constants.rho*constants.eta/2);
    elements.La = constants.rho*L/A;
    elements.Ca = A*L/(constants.rho*constants.c^2);
    elements.Ga = (2*pi*R*L*(constants.nu-1)/...
        (constants.rho*constants.c^2)*sqrt(constants.kappa*w/(2*...
        constants.cp*constants.rho)));
    Rwct = constants.etawc*T/(2*pi*R^3*L);
    Rwst = constants.etaws*T/(2*pi*R^3*L);
    Lwct = constants.rhowc*T/(2*pi*R*L);
    Lwst = constants.rhows*T/(2*pi*R*L);
    %elements.Cwc = 2*pi*R^3*L/(constants.Ewc*T);   % There is a potential
                                                    % typo in Table III, which
                                                    % should perhaps read Cwst
                                                    % rather than Cws
    %elements.Cws = 2*pi*R^3*L/(constants.Ews*T);
```

```
    Cwct = 2*pi*R^3*L/(constants.Ewc*T);
    Cwst = 2*pi*R^3*L/(constants.Ews*T);
    elements.Rwc = Rwct/(C+eps);
    elements.Rws = Rwst/(1-C);
    elements.Lwc = Lwct/(C+eps);
    elements.Lws = Lwst/(1-C);
    elements.Cwc = Cwct*C;
    elements.Cws = Cwst*(1-C);
    elements.Rw = 6500/(2*pi*R*L);
    elements.Lw = 0.4/(2*pi*R*L);
    varargout = {elements};
case 'getDistalLoadImpedance',
    ZL = 0;
    varargout = {ZL};
case 'getZg',
    w = varargin{2};
    glottalDims = varargin{3};
    d = glottalDims{1}; % = 0.01 to 0.2 cm (van den Berg, 1957)
    h = glottalDims{2}; % = 0.32 cm (van den Berg, 1957)
    lg = glottalDims{3}; % = 1.8 cm (van den Berg, 1957)
    Ug = varargin{4};
    mu = 1.8*10^-1;
    K = 0.875;
    rho = 1.14*10^-3;
    R_turbulence_factor = 1;
    Zvf = (12*mu*h/(lg*d^3)+K*R_turbulence_factor*rho*Ug/...
        (2*(lg*d)^2)) + i*w*rho*h/(lg*d); % from Stevens (1998) p. 165.
```

```
    %Zpc = w/eps; % temporary
    %Zg = (Zvf.*Zpc)./(Zvf + Zpc); % impedance along the vocal folds
                                   % (vf) in parallel with the posterior
                                   % glottal chink (pc).
    varargout = {Zvf};
case 'getSGParams',
    p = [
             0   10.0000    0.8000    0.3724    0.6700    1.0000
        1.0000    5.0000    0.6000    0.1735    0.5000    2.0000
        2.0000    2.2000    0.5500    0.1348    0.5000    3.0000
        3.0000    1.1000    0.4000    0.0528    0.3300    3.0000
        4.0000    1.0500    0.3650    0.0409    0.2500    3.0000
        5.0000    1.1300    0.2950    0.0182    0.2000    3.0000
        6.0000    1.1300    0.2950    0.0182    0.0922    3.0000
        7.0000    0.9700    0.2700    0.0168    0.0848    3.0000
        8.0000    1.0800    0.2150    0.0137    0.0669    3.0000
        9.0000    0.9500    0.1750    0.0114    0.0525    3.0000
       10.0000    0.8600    0.1750    0.0114    0.0525    3.0000
       11.0000    0.9900    0.1550    0.0103    0.0449    3.0000
       12.0000    0.8000    0.1450    0.0097    0.0409    3.0000
       13.0000    0.9200    0.1400    0.0094    0.0389    3.0000
       14.0000    0.8200    0.1350    0.0091    0.0369    3.0000
       15.0000    0.8100    0.1250    0.0086    0.0329    3.0000
       16.0000    0.7700    0.1200    0.0083    0.0308    3.0000
       17.0000    0.6400    0.1090    0.0077    0.0262    3.0000
       18.0000    0.6300    0.1000    0.0072    0.0224    3.0000
       19.0000    0.5170    0.0900    0.0066         0    3.0000
```

```
        20.0000    0.4800    0.0800    0.0060         0    3.0000
        21.0000    0.4200    0.0700    0.0055         0    3.0000
        22.0000    0.3600    0.0550    0.0047         0    2.0000
        23.0000    0.3100    0.0480    0.0043         0    2.0000
        24.0000    0.2500    0.0380    0.0038         0    1.0000
        25.0000    0.1100    0.0320    0.0034         0         0
        26.0000    0.1310    0.0270    0.0032         0         0
        27.0000    0.1050    0.0240    0.0031         0         0
        28.0000    0.0750    0.0220    0.0030         0         0
        29.0000    0.0590    0.0400    0.0039         0         0
        30.0000    0.0480    0.0400    0.0039         0         0
        31.0000    0.0480    0.0400    0.0039         0         0
        32.0000    0.0480    0.0400    0.0039         0         0
        33.0000    0.0480    0.0400    0.0039         0         0
        34.0000    0.0480    0.0400    0.0039         0         0
        35.0000    0.0480    0.0400    0.0039         0         0
    ];
    varargout = {p};
case 'symmtrach',
    parameters = varargin{2};
    generations = varargin{3};
    depths = parameters(:,1);
    lengths = parameters(:,2);
    radii = parameters(:,3);
    thickness = parameters(:,4);
    cfrac = parameters(:,5);
    dn = parameters(:,6);
```

```
for j=1:generations
    for k=1:2:2^(j-1),
        if j==1,
            structure(j,:)=0;
            n = structure(j,k) + 1;
            newl(j,k) = lengths(n);
            newr(j,k) = radii(n);
            newh(j,k) = thickness(n);
            newcfrac(j,k) = cfrac(n);
        else
            structure(j,k)=structure(j-1,(k+1)/2)+1;
            n = structure(j,k) + 1;
            newl(j,k) = lengths(n);
            newr(j,k) = radii(n);
            newh(j,k) = thickness(n);
            newcfrac(j,k) = cfrac(n);

            structure(j,k+1)=structure(j-1,(k+1)/2)+1;
            n = structure(j,k+1) + 1;
            newl(j,k+1) = lengths(n);
            newr(j,k+1) = radii(n);
            newh(j,k+1) = thickness(n);
            newcfrac(j,k+1) = cfrac(n);
        end;
    end;
end;
varargout = {newl,newr,newh,newcfrac}; % = {L,R,T,C};
```

```
end;
```

Appendix B

ARCTIC database files which were analyzed in Chapter 4

Table B.1:

CV type	word	file name	word	file name	word	file name
[bi]	be	a0020	be	a0026	be	a0039
	be	a0145	be	a0161	be	a0233
	be	a0255	be	a0262	be	a0270
	be	a0280	be	a0305	be	a0424
	be	a0466	be	a0500	be	a0508
	be	a0524	be	a0559	be	b0009
	be	b0022	be	b0043	be	b0080
	be	b0102	be	b0339	be	b0348
	be	b0362	be	b0443	beach	a0213
	beach	a0554	beach	b0416	beady	a0503
	beating	a0340	beating	b0141		
[bɪ]	big	a0095	big	a0218	big	a0262
	big	a0273	big	a0297	big	a0576

Continued on next page

Table B.1:

CV type	word	file name	word	file name	word	file name
	big	a0581	big	b0145	big	b0156
	big	b0241	big	b0318	big	b0530
	big	a0189	big	a0337	bit	a0431
	bit	b0082	bit	b0434	business	a0250
	business	a0370	business	a0574	business	a0014
[bɛ]	bed	b0417	bed	a0180	beg	b0180
	best	b0538	better	b0017	better	b0191
	better	b0219				
[bæ]	babbling	b0430	back	a0012	back	a0137
	back	a0171	back	a0208	back	a0251
	back	a0335	back	a0387	back	a0508
	back	b0021	back	b0098	back	b0158
	back	b0267	back	b0492	back	b0165
	back	b0329	back	a0411	bad	a0270
	bassett	a0296	bath	a0270		
[bʌ]	buggy	b0344	bustle	b0440	but	a0004
	but	a0072	but	a0103	but	a0127
	but	a0144	but	a0195	but	a0228
	but	a0285	but	a0301	but	a0303
	but	a0358	but	a0360	but	a0361
	but	a0374	but	a0380	but	a0391
	but	a0413	but	a0436	but	a0456
	but	a0462	but	a0466	but	a0515
	but	a0522	but	a0534		
[ba]	Bob	a0387	Bob	a0392	Boston	b0452
	box	a0404				
[bo]	beau	b0156	beau	a0206	both	a0092
	both	a0266	both	a0490	both	a0591

Continued on next page

Table B.1:

CV type	word	file name	word	file name	word	file name
	both	b0112				
[bʊ]	book	b0281	book	b0351	book	b0532
[pi]	peace	a0559	peace	b0455	people	a0024
	people	b0469	people	a0302	people	a0328
	Peterborough	a0569				
[pɪ]	picked	a0113	picture	a0138	picture	b0035
	picture	b0232	pitched	a0546	pits	b0141
[pɛ]	peasant	b0480	pebbles	b0149		
[pæ]	Packard's	b0193	paddling	a0088	pass	a0247
	pass	a0345	pass	a0387	pass	b0162
	pass	b0217	pass	a0064		
[pʌ]	public	a0022	puff	a0073	puzzled	a0219
[pa]	popular	b0057	possible	b0353		
[po]	poked	b0148	postpone	a0578		
[pʊ]	put	a0425	put	b0266	put	b0271
[di]	deed	a0330	deep	a0293	deep	b0169
	deep	a0076	deep	a0096	defect	b0137
	depot	b0269				
[dɪ]	Dick	b0468	did	a0081	did	a0182
	did	a0192	did	a0237	did	a0361
	did	a0377	did	a0526	did	a0555
	did	a0576	did	b0041	did	b0045
	did	b0129	did	b0137	did	b0183
	did	b0261	did	b0263	did	b0275
	did	b0317	did	b0324	did	b0335
	did	b0405	did	b0406	did	b0435
	did	b0443	did	b0444	did	b0507
	did	b0508	did	b0536	did	a0583

Continued on next page

Table B.1:

CV type	word	file name	word	file name	word	file name
	did	a0385	did	a0534	difficulty	b0284
	dig	a0346	dig	b0312	disaffection	b0123
	disappointment	a0429	disconcerting	b0482	disillusionment	a0510
	disinclined	b0223				
[dɛ]	dead	a0087	dead	a0148	dead	a0220
	deaf	b0097	death	a0086	death	a0211
	death	b0212	death	a0528	death	a0544
	death	b0383	debutante	a0120		
[dæ]	daddy	a0379	dashed	b0233		
[dʌ]	double	a0072	double	a0318	double	a0261
	dozen	a0524	ducks	b0323		
[da]	doctor	b0310	doctrine	a0535	document	a0566
	dog	a0197	dog	b0239	dog	b0466
	dog	a0585	dog	b0152	dog	b0466
	dog	a0196	dog	a0341	dog	a0405
	dog	b0007				
[du]	do	a0025	do	a0368	do	b0001
	do	b0069	do	b0231	do	b0343
	do	b0462	duplicity	b0226	duty	b0509
[ti]	tea	b0252	teach	b0367	teach	b0369
	tee	a0388	teeth	b0159	teeth	a0279
[tɛ]	test	a0509	test	a0408		
[tæ]	taboo	b0210	taboo	a0433	taboo	b0510
	tacit	a0298	tap	a0465	tap	a0153
	tap	a0529	task	b0481		
[tʌ]	touchy	b0209				
[ta]	talk	b0122	talk	b0214	talk	b0518
	talk	b0539	talk	a0043	taught	a0531

Continued on next page

Table B.1:

CV type	word	file name	word	file name	word	file name
	taught	b0366	taut	a0498		
[tʊ]	took	a0494	took	a0497	took	b0011
	took	b0039	took	b0046	took	b0143
	took	b0476	took	b0490		
[tu]	too	a0472	too	a0472	too	b0504
	too	b0036	too	a0467	tooth	a0285
	Tudor	b0222	Tudor	b0255	Tudor	b0257
	Tuesday	b0391	two	a0111	two	a0121
	two	a0122	two	a0166	two	a0188
	two	a0210	two	a0371	two	a0415
	two	b0115	two	b0124	two	b0164
	two	b0236	two	b0277	two	a0184
	two	b0048	two	b0060	two	b0365
[gi]	geese	b0323				
[gɪ]	give	a0181	give	a0271	give	a0278
	give	a0286	give	a0513	give	b0056
	give	b0247	give	a0347	give	b0379
[gɛ]	get	a0346	get	b0038	get	b0382
	guess	b0462	guess	b0210		
[gæ]	gad	a0008	gad	b0001		
[ga]	god	a0006	god	a0141	god	b0185
	gosh	b0339	got	a0264	got	a0321
	got	b0065	got	b0195	got	b0231
	got	b0296				
[gʊ]	good	a0230	good	a0569	good	a0574
	good	a0593	good	b0207	good	b0215
	good	b0276	good	b0390	good	b0517
	good	b0312	good	a0158	good	a0150

Continued on next page

Table B.1:

CV type	word	file name	word	file name	word	file name
	good	a0556				
[gu]	goose	a0547	goose	b0502		
[ki]	keep	a0101	keep	a0287	keep	a0394
	keep	b0060	key	b0028		
[kɪ]	kiddies	a0275				
[kæ]	cabin	a0448	cabin	b0051	cabin	a0167
	cabin	a0049	cabin	a0277	cabin	b0050
	capital	a0023	capital	a0364	captain	a0264
	captain	a0466	captain	b0374	captain	b0526
	captain	b0528	captured	a0290	cascades	a0464
	cash	a0374	cash	a0593	castor	b0173
	casual	b0271	casualty	a0292	catch	a0055
	catch	a0356				
[kʌ]	couple	b0046	cut	a0269	cutters	a0375
[kɑ]	caught	a0200	caught	b0032	caught	b0155
	caused	b0480	caution	b0075	cautiously	a0194
	Cocky	a0582	cod	a0234	copper	b0425
	cottonwoods	b0204				
[ko]	coat	a0494	code	b0500		
[kʊ]	cook	b0374	cooking	b0433	could	a0107
	could	a0295	could	a0409	could	a0424
	could	a0457	could	a0521	could	a0523
	could	b0069	could	b0075	could	b0076
	could	b0123	could	b0242	could	b0270
	could	b0375	could	b0400	could	b0464
	could	b0485	could	a0534		

Bibliography

[1] T. V. Ananthapadmanabha and G. Fant. Calculation of true glottal flow and its components. *Speech Communication*, 1:167–184, 1982.

[2] Sylvain Bromberger and Morris Halle. The contents of phonological signs: A comparison between their use in derivational theories and optimality theories. In Iggy Roca, editor, *Derivations and Constraints in Phonology*, pages 93–123. Clarendon Press, Oxford, 1997.

[3] Harold A. Cheyne. *Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck*. PhD thesis, MIT, 2002.

[4] Xuemin Chi. The quantal effect of sub-glottal resonance on vowel formant. RQE report, MIT, 2005.

[5] Xuemin Chi and Morgan Sonderegger. Subglottal coupling and vowel space. *The Journal of the Acoustical Society of America*, 115(5):2540–2540, 2004.

[6] Tsutomu Chiba and Masato Kajiyama. *The Vowel: Its structure and nature*. Phonetic Society of Japan, Tokyo, 1941/1958.

[7] Ludmilla A. Chistovich. Central auditory processing of peripheral vowel spectra. *The Journal of the Acoustical Society of America*, 77(3):789–805, 1985.

[8] Noam Chomsky. Langauge as a natural object. In *New Horizons in the Study of Language and Mind*.

[9] Noam Chomsky and Morris Halle. *The Sound Pattern of English*. MIT Press, Cambridge, MA, 1968.

[10] Bert Cranen and Louis Boves. On subglottal formant analysis. *The Journal of the Acoustical Society of America*, 81(3):734–746, 1987.

[11] Pierre C. Delattre, Alvin M. Liberman, and Franklin S. Cooper. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, 27(4):769–773, 1955.

[12] Jerold A. Edmondson and John H. Esling. The valves of the throat and their functioning in tone, vocal register, and stress: laryngoscopic case studies. *Phonology*, 2006.

[13] Gunnar Fant. *Acoustic Theory of Speech Production: With calculations based on X-Ray studies of Russian articulations*. Mouton, The Hague, 1960.

[14] Gunnar Fant, Kenzo Ishizaka, J. Lindqvist, and J. Sundberg. Subglottal formants. *STL-QPSR*, 1:1–12, 1972.

[15] James L. Flanagan. Difference limen for formant amplitude. *Journal of Speech and Hearing Disorders*, 22(2):205–212, 1957.

[16] James L. Flanagan. *Speech Analysis, Synthesis and Perception*. Springer-Verlag, Berlin, second edition, 1972.

[17] Jeffrey J. Fredberg and A. Hoenig. Mechanical response of the lungs at high frequencies. *Journal of Biomechanical Engineering*, 100:57–66, 1978.

[18] David Fruchter and Harvey M. Sussman. The perceptual relevance of locus equations. *Journal of the Acoustical Society of America*, 102(5):2997–3008, 1997.

[19] Robert H. Habib, Richard B. Chalker, Béla Suki, and Andrew C. Jackson. Airway geometry and wall mechanical properties estimated from subglottal input impedance in humans. *Journal of Applied Physiology*, 77(1):441–451, 1994.

[20] Morris Halle. Feature geometry and feature spreading. *Linguistic Inquiry*, 26:1–46, 1995.

[21] Helen M. Hanson. Measurements of subglottal resonances and their influence on vowel spectra. *The Journal of the Acoustical Society of America*, 100(4):2656–2656, 1996.

[22] V. Paul Harper, Steven S. Kraman, Hans Pasterkamp, and George R. Wodicka. An acoustic model of the respiratory tract. *IEEE Transactions on Biomedical Engineering*, 48(5):543–550, 2001.

[23] V. Paul Harper, Hans Pasterkamp, Hiroshi Kiyokawa, and George R. Wodicka. Modeling and measurement of flow effects on tracheal sounds. *IEEE Transactions on Biomedical Engineering*, 50(1):1–10, 2003.

[24] Keith Horsfield, Gladys Dart, Dan E. Olson, Giles F. Filly, and Gordon Cumming. Models of the human bronchial tree. *Journal of Applied Physiology*, 31(2):207–217, 1971.

[25] Herbert Hudde and Harald Slatky. The acoustical input impedance of excised human lungs—measurements and model matching. *The Journal of the Acoustical Society of America*, 86(2):475–492, 1989.

[26] K. Ishizaka, M. Matsudaira, and T. Kaneko. Input acoustic-impedance measurement of the subglottal system. *The Journal of the Acoustical Society of America*, 60(1):190–197, 1976.

[27] Masashi Ito, Jun Tsuchida, and Masafumi Yano. On the effectiveness of whole spectral shape for vowel perception. *The Journal of the Acoustical Society of America*, 110(2):1141–1149, 2001.

[28] Andrew C. Jackson, Béla Suki, Melih Ucar, and Robert H. Habib. Branching airway network models for analyzing high-frequency lung input impedance.

[29] Roman Jakobson. *Selected Writings*, volume I. Mouton de Gruyter, Berlin, second edition, 1971.

[30] Roman Jakobson, Gunnar Fant, and Morris Halle. *Preliminaries to Speech Analysis*. MIT Press, Cambridge, MA, 1962.

[31] Ken J. Kallail and Floyd W. Emanuel. An acoustic comparison of isolated whispered and phonated vowel samples produced by adult male subjects. *Journal of Phonetics*, 12:175–186, 1984.

[32] Ken J. Kallail and Floyd W. Emanuel. Formant-frequency differences between isolated whispered and phonated vowel samples produced by adult female subjects. *Journal of Speech and Hearing Research*, 27:245–251, 1984.

[33] P. A. Keating. The window model of coarticulation: articulatory evidence. In J. Kingston and M. Beckman, editors, *Papers in Laboratory Phonology*, volume I, pages 451–470. Cambridge University Press, Cambridge, UK, 1990.

[34] Samuel J. Keyser. Enhancement revisited. In Michael Kenstowicz, editor, *Ken Hale: A life in language*, pages 271–291. MIT Press, Cambridge, MA, 2001.

[35] Samuel J. Keyser and Kenneth N. Stevens. Enhancement and overlap in the speech chain. *Language*, 82(1):33–63, 2006.

[36] Michael Kiefte and Keith R. Kluender. The relative importance of spectral tilt in monophthongs and diphthongs. *The Journal of the Acoustical Society of America*, 117(3):1395–1404, 2005.

[37] Dennis H. Klatt and Laura C. Klatt. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America*, 87(2):820–857, 1990.

[38] John Kominek and Alan W. Black. CMU ARCTIC databases for speech synthesis. Technical Report CMU-LTI-03-177, Language Technologies Institute, School of Computer Science, 2003. Carnegie Mellon University.

[39] B. Lindblom. On vowel reduction. (29), 1963. Speech Transmission Laboratory.

[40] Karen Livescu. *Feature-based pronunciation modeling for automatic speech recognition*. PhD thesis, MIT, 2005.

[41] C. I. Malme. Detectability of small irregularities in a broadband noise spectrum. *RLE Quarterly Progress Report*, 52:139–142, 1959.

[42] C. I. Malme. Detectability of small irregularities in a harmonic line spectrum. *RLE Quarterly Progress Report*, 57:122–126, 1960.

[43] Philip M. Morse. *Vibration and Sound*. Acoustical Society of America, Washington, D.C., 1936/1981.

[44] F. Rohrer. Der strömmungswiderstand in den menschlichen atemwegen und der einfluss der unregelmässigen verzweigung des bronchialsystems auf den atmungsverlauf in verschiedenen lungbezirken. *Pflügers Archiv*, 162:225–299, 1915.

[45] Morgan Sonderegger. Sublottal coupling and vowel space: An investigation in quantal theory. Undergraduate thesis, MIT, 2004.

[46] Kenneth N. Stevens. The quantal nature of speech: Evidence from articulatory-acoustic data. In P. B. Denes and Jr. E. E. David, editors, *Human Communication, A Unified View*. McGraw-Hill, New York, 1972.

[47] Kenneth N. Stevens. On the quantal nature of speech. *Journal of Phonetics*, 17:3–45, 1989.

[48] Kenneth N. Stevens. *Acoustic Phonetics*. MIT Press, Cambridge, MA, 1998.

[49] Kenneth N. Stevens. Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America*, 111(4):1872–1891, 2002.

[50] Kenneth N. Stevens and Helen M. Hanson. Classification of glottal vibration from acoustic measurements. In O. Fujimura and M. Hirano, editors, *Vocal Fold Physiology: Voice quality control*, pages 147–170. Singular, San Diego, 1995.

[51] Brad H. Story. Synergistic modes of vocal tract articulation for english vowels. *Journal of the Acoustical Society of America*, 118(6):3834–3859, 2005.

[52] Harvey M. Sussman, Nicola Bessell, Eileen Dalston, and Tivoli Majors. An investigation of stop place of articulation as a function of syllable position: A locus equation perspective. *Journal of the Acoustical Society of America*, 101(5):2826–2838, 1997.

[53] Ann K. Syrdal and H. S. Gopal. A perceptual model of vowel recognition based on the auditory representation of american english vowels. *The Journal of the Acoustical Society of America*, 79(4):1086–1100, 1986.

[54] Nikolaj S. Trubetzkoi. *Grundzge der Phonology.* Vandenhoeck und Ruprecht, Gttingen, 1939/1958.

[55] Janwillem van den Berg. An electrical analogue of the trachea, lungs and tissues. *Acta Physiol. Pharmacol. Neerlandica.*, 9:361–385, 1960.

[56] Janwillem van den Berg, J. T. Zantema, and Jr. P. Doornenbal. On the air resistance and the bernoulli effect of the human larynx. *Journal of the Acoustical Society of America*, 29(5):626–631, 1957.

[57] E. R. Weibel. *Morphometry of the Human Lung.* Springer, Berlin, 1963.

[58] Robert Willis. On vowel sounds, and on reed-organ pipes. *Trans. Camb. Phil. Soc.*, 3:231–263, 1829.